THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Iterative multilevel particle approximation for McKean–Vlasov SDEs

Citation for published version:
Szpruch, L, Tan, S & Tse, A 2019, 'Iterative multilevel particle approximation for McKean–Vlasov SDEs', *Annals of Applied Probability*, vol. 29, no. 4, pp. 2230-2265. https://doi.org/10.1214/18-AAP1452

Digital Object Identifier (DOI):
10.1214/18-AAP1452

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Annals of Applied Probability

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 07. May. 2023

arXiv:1706.00907v2 [math.PR] 16 Jan 2019

Iterative Multilevel Particle Approximation for McKean-Vlasov SDEs

Lukasz Szpruch[1,2], Shuren Tan[1], and Alvin Tse[1]

[1]School of Mathematics,University of Edinburgh
[2]The Alan Turing Institute, London

Abstract

The mean field limits of systems of interacting diffusions (also called stochastic interacting particle systems (SIPS)) have been intensively studied since McKean [25] as they pave a way to probabilistic representations for many important nonlinear/nonlocal PDEs. The fact that particles are not independent render classical variance reduction techniques not directly applicable and consequently make simulations of interacting diffusions prohibitive.

In this article, we provide an alternative iterative particle representation, inspired by the fixed point argument by Sznitman [30]. The representation enjoys suitable conditional independence property that is leveraged in our analysis. We establish weak convergence of iterative particle system to the McKean-Vlasov SDEs (McKV-SDEs). One of the immediate advantages of iterative particle system is that it can be combined with the Multilevel Monte Carlo (MLMC) approach for the simulation of McKV-SDEs. We proved that the MLMC approach reduces the computational complexity of calculating expectations by an order of magnitude. Another perspective on this work is that we analyse the error of nested Multilevel Monte Carlo estimators, which is of independent interest. Furthermore, we work with state dependent functionals, unlike scalar outputs which are common in literature on MLMC. The error analysis is carried out in uniform, and what seems to be new, weighted norms.

2010 AMS subject classifications: Primary: 65C30, 60H35; secondary: 60H30.

Keywords : Mckean-Vlasov SDEs, Stochastic Interacting Particle Systems, Non-linear Fokker-Planck equations, Probabilistic Numerical Analysis

1 Introduction

The theory of mean field interacting particle systems was pioneered by the work of H. McKean [25], where he gave a probabilistic interpretation of a class of nonlinear (due to the dependence on the coefficients of the solution itself) nonlocal PDEs. Probabilistic representation has an advantage, as it paves a way to Monte-Carlo approximation methods which are efficient in high dimensions. Fix $T > 0$. Let $\{W_t\}_{t\in[0,T]}$ be an r-dimensional Brownian motion on a filtered probability space $(\Omega, \{\mathcal{F}_t\}_t, \mathcal{F}, \mathbb{P})$. Consider continuous functions $b : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d\otimes r}$ and their corresponding non-linear (in the sense of McKean) stochastic differential equation (McKV-SDE) given by

$$\begin{cases} dX_t &= b[X_t, \mu_t^X]\, dt + \sigma[X_t, \mu_t^X]\, dW_t, \\ \mu_t^X &= \mathrm{Law}(X_t), \quad t \in [0,T], \end{cases} \tag{1.1}$$

where $X_0 \sim \mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ and $G[x,m] := \int_{\mathbb{R}^d} G(x,y)\, m(dy)$, for any $x \in \mathbb{R}^d$ and $m \in \mathcal{P}_2(\mathbb{R}^d)$ (square-integrable laws on $\mathbb{R}^d$). Notice that $\{X_t\}_{t\in[0,T]}$ is not necessarily a Markov process and hence it is not immediate what the corresponding backward Kolmogorov equation looks like. Nonetheless using Itô's formula with $P \in C_b^2(\mathbb{R}^d)$, one can derive corresponding nonlinear Kolmogorov-Fokker-Planck equation

$$\partial_t \langle \mu_t, P\rangle = \langle \mu_t, \frac{1}{2}\sum_{i,j=1}^{d} \partial^2_{x_i,x_j} P(\cdot)\big(\sigma\sigma^T\big)_{ij}[\cdot, \mu_t] + \sum_{i=1}^{d} \partial_{x_i} P(\cdot) b_i[\cdot, \mu_t]\rangle, \tag{1.2}$$

where $\langle m, F\rangle := \int_{\mathbb{R}^d} F(y)\, m(dy)$, [2, 11, 30]. The theory of propagation of chaos, [30], states that (1.1) arises as a limiting equation of the system of interacting diffusions $\{Y_t^{i,N}\}_{i=1,\dots,N}$ on $(\mathbb{R}^d)^N$ given by

$$\begin{cases} dY_t^{i,N} &= b[Y_t^{i,N}, \mu_t^{Y,N}]dt + \sigma[Y_t^{i,N}, \mu_t^{Y,N}]dW_t^i, \\ \mu_t^{Y,N} &:= \frac{1}{N}\sum_{i=1}^{N} \delta_{Y_t^{i,N}}, \quad t \geq 0, \end{cases} \tag{1.3}$$

where $\{Y_0^{i,N}\}_{i=1,\dots,N}$ are i.i.d samples with law μ_0 and $\{W_t^i\}_{i=1,\dots,N}$ are independent Brownian motions. It can be shown, under sufficient regularity conditions on the coefficients, that $\mu^{Y,N} \in \mathcal{P}_2(C([0,T], \mathbb{R}^d))$ converges in law to μ^X, see [26]. This is a not trivial result as the particles are not independent. Moreover, (1.3) can be interpreted as a first step towards numerical schemes for (1.1). To obtain a fully implementable algorithm one needs to study time discretisation of (1.1). As in seminal papers by Bossy and Talay [6, 7] we work with an Euler scheme. Take partition $\{t_k\}_k$ of $[0,T]$, with $t_k - t_{k-1} = h$ and define $\eta(t) := t_k$ if $t \in [t_k, t_{k+1})$. The continuous Euler scheme reads

$$\overline{Y}_t^{i,N} = \overline{Y}_{t_k}^{i,N} + b[\overline{Y}_{\eta(t)}^{i,N}, \mu_{\eta(t)}^{\overline{Y},N}](t - t_k) + \sigma[\overline{Y}_{\eta(t)}^{i,N}, \mu_{\eta(t)}^{\overline{Y},N}](W_t^i - W_{t_k}^i)\,. \tag{1.4}$$

Note that due to interactions between discretised diffusions, implementation of (1.4) requires N^2 arithmetic operations at each step t_k of the scheme. This makes simulations of (1.4) very costly, but should not come as a surprise as the aim is to approximate non linear/non local PDEs (1.2) for which the deterministic schemes based on space discretisation, typically, are also computationally very demanding [4]. It has been proven that the empirical distribution function of N particles (1.4) converges, in a weak sense, to the distribution of the corresponding McKean-Vlasov limiting equation with the rate $O((\sqrt{N})^{-1} + h)$, see [2, 3, 5, 7]. Hence the computational cost of achieving a mean-square-error (see Theorem 4.6 for the definition) of order $\epsilon^2 > 0$ using this direct approach is $\mathcal{O}(\epsilon^{-5})$.

The lack of independence among interacting diffusions and the fact that the statistical error coming from approximating a measure creates a bias in the approximation, render applications of variance reduction techniques non-trivial. In fact, we are not aware of any rigorous work on variance reduction techniques for McKV-SDEs. In this article, we develop an iterated particle system that allows decomposing the statistical error and bias. We also provide an error analysis for a general class of McKV-SDEs. Finally, we deploy the MLMC method of Giles-Heinrich [16, 19] (see also 2-level MC of Kebaier [20]). In Section 2.2, we show that a direct application of MLMC to (1.3) fails. It is worth pointing out that the idea of combining an iterative method with MLMC to solve non-linear PDEs has very recently been proposed in [14]. However, their interest is on BSDEs and their connections to semi-linear PDEs.

The key technical part of the paper is weak convergence analysis of the time discretisation that allows for iteration of the error in a suitable norms. It is well know, at least since the work [31] that weak error analysis relies on the corresponding PDE theory. However as we already stated the solution to (1.1) is not Markovian on $\mathbb{R}^d$. To overcome we work with forward backward system

$$\begin{cases} X_t^{0,X_0} &= \xi + \int_0^t b[X_s^{s,\xi}, \mu_s^{X^{0,\xi}}]\, ds + \int_0^t \sigma[X_s^{0,X_0}, \mu_s^{X^{0,X_0}}]\, dW_s, \\ \mu_t^{X^{0,X_0}} &= \text{Law}(X_t^{0,X_0}), \end{cases}$$

and note that $X_t^{0,X_0} \neq X_t^{0,x}|_{x=X_0}$ in general (see [8]). This makes building of standard PDE theory on $[0,T] \times \mathbb{R}^d$ problematic and lead to theory of PDEs on measure spaces proposed by P. Lions in his lectures in Collège de France ([24]) and further developed in [8, 11]. Here we work with

$$\mathcal{X}_t^{0,x} = x + \int_0^t b[\mathcal{X}_s^{0,x}, \mu_s^{X^{0,\xi}}]\, ds + \int_0^t \sigma[\mathcal{X}_s^{0,x}, \mu_s^{X^{0,\xi}}] dW_s. \tag{1.5}$$

Notice that (1.5), unlike (1.1), is a Markov process. Furthermore, if (1.1) has a unique (weak) solution, then $\mathcal{X}_t^{0,x}|_{x=X_0} = X_t^{0,X_0}$. This means that

$$\int_{\mathbb{R}^d} \mathbb{E}\big[P(\mathcal{X}_t^{0,x})\big]\, \mu_0(dx) = \mathbb{E}\big[\mathbb{E}[P(X_t)|X_0]\big].$$

It can be shown that $v(0,x) = \mathbb{E}\big[P(\mathcal{X}_t^{0,x})\big]$ is a solution to backward Kolmogorov equation on $[0,T] \times \mathbb{R}^d$ which we will explore in this paper.

1.1 Iterated particle method

The main idea is to approximate (1.1) with a sequence of classical SDEs defined as

$$dX_t^m = b[X_t^m, \mu_t^{X^{m-1}}]dt + \sigma[X_t^m, \mu_t^{X^{m-1}}]dW_t^m, \quad \mu_0^{X^m} = \mu_0^X, \tag{1.6}$$

where (W^m, X_0^m) are independent for all $m \in \mathbb{N}$ as well as (W^m, X_0^m) and (W^n, X_0^n) $m \neq n \in \mathbb{N}$, are independent. The conditional independence across iterations is the key difference of our approach from the proof of existence of solutions by Sznitman [30], where the same Brownian motion and initial condition are used at every iteration. The Euler scheme with $\mu_0^{\overline{X}^m} = \mu_0^X$ reads

$$d\overline{X}_t^m = b[\overline{X}_{\eta(t)}^m, \mu_{\eta(t)}^{\overline{X}^{m-1}}]dt + \sigma[\overline{X}_{\eta(t)}^m, \mu_{\eta(t)}^{\overline{X}^{m-1}}]dW_t^m\,. \tag{1.7}$$

To implement (1.7) at every step of the scheme, one needs to compute the integral with respect to the measure from the previous iteration $m-1$. This integral is calculated by approximating measure $\mu_{\eta(t)}^{\overline{X}^{m-1}}$ by the empirical measure with N_{m-1} samples. Consequently, we take $\mu_0^{\overline{Y}^{i,m}} = \mu_0^X$ and define, for $m \in \mathbb{N}$ and $1 \leq i \leq N_m$,

$$d\overline{Y}_t^{i,m} = b[\overline{Y}_{\eta(t)}^{i,m}, \mu_{\eta(t)}^{\overline{Y}^{m-1},N_{m-1}}]dt + \sigma[\overline{Y}_{\eta(t)}^{i,m}, \mu_{\eta(t)}^{\overline{Y}^{m-1},N_{m-1}}]dW_t^{i,m}, \tag{1.8}$$

and call it an *iterative particle system*. As above, we require that $W^{i,m}$, $1 \leq i \leq N_m$, $m \in \mathbb{N}$, and $\overline{Y}_0^{i,m}$, $1 \leq i \leq N_m$, $m \in \mathbb{N}$, are independent. By this construction, the particles $(\overline{Y}_t^{i,m})_i$ are independent upon conditioning on $\sigma\Big(\{\overline{Y}_t^{i,m-1}\}_{1\leq i\leq N_{m-1}} : t \in [0,T]\Big)$. The error analysis of (1.8) is presented in Theorem (4.6) and (4.7). From there one can deduce that optimal computational cost is achieved when $\{N_m\}_m$ is increasing and the computational complexity of computing expectations with (1.8) is of the same order as the original particle system, i.e. ϵ^{-5}.

1.2 Main result of the iterative MLMC algorithm

To reduce the computational cost, we combine the MLMC method with Picard iteration (1.6). Fix m and L. Let $\Pi^\ell = \{0 = t_0^\ell, \dots, t_k^\ell, \dots, T = t_{2^\ell}^\ell\}$, $\ell = 0, \dots, L$, be a family of time grids such that $t_k^\ell - t_{k-1}^\ell = h_\ell = T2^{-\ell}$. To simulate (1.7) at Picard step m and for all discretisation levels ℓ we need to have an approximation of the relevant expectations with respect to the law of the process at the previous Picard step $m-1$ and the time grid Π^L, i.e.

$$\Big(\mathbb{E}[b(x, \overline{X}_0^{m-1})], \dots, \mathbb{E}[b(x, \overline{X}_{t_k^L}^{m-1})], \dots, \mathbb{E}[b(x, \overline{X}_T^{m-1})]\Big)\,,$$
$$\Big(\mathbb{E}[\sigma(x, \overline{X}_0^{m-1})], \dots, \mathbb{E}[\sigma(x, \overline{X}_{t_k^L}^{m-1})], \dots, \mathbb{E}[\sigma(x, \overline{X}_T^{m-1})]\Big)\,.$$

By approximating these expectations with the MLMC (signed) measure $\mathcal{M}^{(m-1)}$ (see Section 2.3 for its exact definition), we arrive at the *iterative MLMC particle* method defined as

$$dY_t^{i,m,\ell} = \langle \mathcal{M}^{(m-1)}_{\eta_\ell(t)}, b(Y^{i,m,\ell}_{\eta_\ell(t)}, \cdot)\rangle \, dt + \langle \mathcal{M}^{(m-1)}_{\eta_\ell(t)}, \sigma(Y^{i,m,\ell}_{\eta_\ell(t)}, \cdot)\rangle \, dW_t^{i,m} \,, \tag{1.9}$$

where $Y^{i,0,\ell} = X_0$. Under the assumptions listed in Section 2, the main result of this paper gives precise error bounds for (1.9).

Theorem 1.1. *Assume* **(Ker-Reg)** *and* **(μ_0-L_p)** *. Fix $M > 0$ and let $P \in C_b^2(\mathbb{R}^d)$. Define $MSE_t^{(M)}(P) := \mathbb{E}[(\langle \mathcal{M}_t^{(M)}, P\rangle - \mathbb{E}[P(X_t)])^2]$. Then there exists a constant $c > 0$ (independent of the choices of M, L and $\{N_{m,\ell}\}_{m,\ell}$) such that for every $t \in [0,T]$,*

$$MSE^{(M)}_{\eta_L(t)}(P) \le c\Big\{ h_L^2 + \sum_{m=1}^{M} \frac{c^{M-m}}{(M-m)!} \cdot \sum_{\ell=0}^{L} \frac{h_\ell}{N_{m,\ell}} + \frac{c^{M-1}}{M!} \Big\}.$$

The proof can be found in Section 4.2. The first term in the above error comes from the analysis of weak convergence for the Euler scheme. The second contains the usual MLMC variance and shows that computational effort should be increasing with with iteration m (rather than equally distributed across iterations). Finally the last term is an extra error due to iterations. Using this result, we prove in Theorem 4.5 that the overall complexity of the algorithm is of order $\epsilon^{-4}|\log \epsilon|^3$ (i.e. one order of magnitude better than the direct approach). We remark that the MLMC measure acts on functionals that depend on spatial variables. We work with uniform norms as in [19, 17], but also introduce suitable weighted norms, which seems new in MLMC literature.

We remark that, the analysis of stochastic particles systems is of independent interest, as it is used as models in molecular dynamics; physical particles in fluid dynamics [28]; behaviour of interacting agents in economics or social networks [10] or interacting neurons in biology [13]. It is also used in modelling networks of neurons (see [12]) and modelling altruism (see [14]).

1.3 Convention of notations

We use $\|A\|$ to denote the Hilbert-Schmidt norm while $|\mathbf{v}|$ is used to denote the Euclidean norm. For any stochastic process $R = \{R_t\}_{t\in I}$, the law of R_t at any time point $t \in I$ is denoted by μ_t^R. $\mathcal{P}_2(E)$ denotes the set of square-integrable probability measures on any Polish space E. On the other hand, $\mathcal{P}_2^s(E)$ denotes, on any Polish space E, the set of random signed measures that are square-integrable almost surely.

Moreover, we denote by $C^{0,2}_{b,p}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R})$ the set of functions P from $\mathbb{R}^m \times \mathbb{R}^n$ to $\mathbb{R}$ that are continuously twice-differentiable in the second argument, for which there exists a constant L such that for each $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, $i, j \in \{1, \dots, n\}$,

$$|\partial_{y_i} P(x,y)| \le L(1+|y|^p), \qquad |\partial^2_{y_i,y_j} P(x,y)| \le L(1+|y|^p),$$

where ∂_{y_i} and $\partial^2_{y_i,y_j}$ denote respectively the first and second order partial derivatives w.r.t. the second argument. Finally, we denote by $C^{p,q}_{b,b}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R})$ the set of functions from $\mathbb{R}^m \times \mathbb{R}^n$ to $\mathbb{R}$ that are continuously p times differentiable in the first argument and continuously q times differentiable in the second argument such that the partial derivatives (up to the respective orders, excluding the "zeroth" order derivative) are bounded.

2 The iterative MLMC algorithm

2.1 Main assumptions on the McKean-Vlasov SDE

Here we state the assumptions needed for the analysis of equation (1.1).

Assumption 2.1.

(Ker-Reg) *The kernels b and σ belong to the sets $C^{2,1}_{b,b}(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^d) \cap C^{0,2}_{b,p}(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^d)$ and $C^{2,1}_{b,b}(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^{d\otimes r}) \cap C^{0,2}_{b,p}(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^{d\otimes r})$ respectively.*

(μ_0-L_p) *The initial law $\mu_0 := \mu_0^X$ satisfies the following condition: for any $p \geq 1$, $\mu_0 \in L^p(\Omega; \mathbb{R}^d)$, i.e.*

$$\int_{\mathbb{R}^d} |x|^p \mu_0(dx) < \infty.$$

Note that if **(Ker-Reg)** holds, then

(Lip) the kernels b and σ are globally Lipschitz, i.e. for all $x_1, x_2, y_1, y_2 \in \mathbb{R}^d$, there exists a constant L such that

$$|b(x_1, y_1) - b(x_2, y_2)| + \|\sigma(x_1, y_1) - \sigma(x_2, y_2)\| \leq L(|x_1 - x_2| + |y_1 - y_2|).$$

If **(Ker-Reg)** and **(μ_0-L_p)** hold, then a weak solution to (1.1) exists and pathwise uniqueness holds (see [30]). In other words $\{X_t\}_{t\geq 0}$ induces a unique probability measure on $C([0,T], \mathbb{R}^d)$. Furthermore it has a property that

$$\sup_{0\leq t\leq T} \mathbb{E}|X_t|^p < \infty. \tag{2.1}$$

The additional smoothness stipulated in **(Ker-Reg)** is needed in the analysis of weak approximation errors.

2.2 Direct application of MLMC to interacting diffusions

There are two issues pertaining to the direct application of MLMC methodology to (1.4): i) the telescopic property needed for MLMC identity [16] does not hold in general; ii) a small number of simulations (particles) on fine time steps (a reason for the improved computational cost in MLMC setting) would lead to a poor approximation of the measure, leading to a high bias. To show that telescopic sum does not hold in general, consider a collection of discretisations of $[0, T]$ with different resolutions. To this end, we fix $L \in \mathbb{N}$. Then Y_T^{i,ℓ,N_ℓ}, $\ell = 1, \dots, L$, denotes for each i a particle corresponding to (1.4) with time-step h_ℓ, where N_ℓ is the total number of particles. Let $P : \mathbb{R}^d \to \mathbb{R}$ be any Borel-measurable function. With a direct application of MLMC in time for (1.4), we replace the standard Monte-Carlo estimator on the left-hand side by an MLMC estimator on the right-hand side as follows.

$$
\begin{aligned}
&\frac{1}{N_L}\sum_{i=1}^{N_L} P(Y_t^{i,L,N_L}) \\
\approx\ & \frac{1}{N_0}\sum_{i=1}^{N_0} P(Y_t^{i,0,N_0}) + \sum_{\ell=0}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\Big[P(Y_t^{i,\ell,N_\ell}) - P(Y_t^{i,\ell-1,N_\ell})\Big].
\end{aligned} \tag{2.2}
$$

However, we observe that such a direct application is not possible, since, in general,

$$
\mathbb{E}\Big[P(Y_t^{1,\ell,N_\ell})\Big] \neq \mathbb{E}\Big[P(Y_t^{1,\ell,N_{\ell+1}})\Big],
$$

which means that we do not have equality in expectation on both sides of (2.2). On the contrary, if we required the number of particles for all the levels to be the same, then the telescopic sum would hold, but clearly, there would be no computational gain from doing MLMC. We are aware of two articles that tackle the aforementioned issue. The case of linear coefficients is treated in [29], in which particles from all levels are used to approximate the mean field at the final (most accurate) approximation level. It is not clear how this approach could be extended to general McKean-Vlasov equations. A numerical study of a "multi-cloud" approach is presented in [18]. The algorithm resembles the MLMC approach to the nested simulation problem in [1, 17, 9, 23]. Their approach is very natural, but because particles within each cloud are not independent, one faces similar challenges as with the classical particle system.

2.3 Construction of the iterative MLMC algorithm

We approximate each of the expectations by the MLMC method, but only have access to samples at grid points Π^ℓ that correspond to $(Y^{i,m-1,\ell})_{i,\ell}$. Consequently, for $\ell < \ell'$, the empirical measure $\frac{1}{N}\sum_{i=1}^{N}\delta_{Y_t^{i,m-1,\ell}}$ is only defined at every timepoint in Π^ℓ, but not $\Pi^{\ell'}$ and one cannot build MLMC telescopic sum across all discretisation levels. For that reason (as in original development of MLMC by Heinrich [19]), we introduce a linear-interpolated measure (in time) $\widetilde{\mu}_t^{Y^{m-1,\ell},N}$ given by

$$\widetilde{\mu}_t^{Y^{m-1,\ell},N} := \begin{cases} \frac{1}{N}\sum_{i=1}^N \delta_{Y_t^{i,m-1,\ell}} & , t \in \Pi^\ell, \\ \\ \left[\frac{t-\eta_\ell(t)}{h_\ell}\right]\widetilde{\mu}_{\eta_\ell(t)+h_\ell}^{Y^{m-1,\ell},N} + \left[1-\frac{t-\eta_\ell(t)}{h_\ell}\right]\widetilde{\mu}_{\eta_\ell(t)}^{Y^{m-1,\ell},N} & , t \notin \Pi^\ell, \end{cases} \tag{2.3}$$

where $\eta_\ell(t) := t_k^\ell, \quad$ if $t \in [t_k^\ell, t_{k+1}^\ell)$. For any continuous function $P : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and any $x \in \mathbb{R}^d$, we define the MLMC signed measure $\mathcal{M}_t^{(m-1)}$ by

$$\langle \mathcal{M}_t^{m-1}, P(x,\cdot)\rangle := \langle \sum_{\ell=0}^L (\widetilde{\mu}_t^{Y^{m-1,\ell},N_{m-1,\ell}} - \widetilde{\mu}_t^{Y^{m-1,\ell-1},N_{m-1,\ell}}), P(x,\cdot)\rangle, \tag{2.4}$$

where $\widetilde{\mu}_t^{Y^{m-1,-1},N_{m,0}} := 0$. We interpret the MLMC operator in a componentwise sense. We then define the particle system $\{Y^{i,m,\ell}\}$ as in (1.9). As usual for MLMC estimators, at each level ℓ, we use the same Brownian motion to simulate particle systems $(Y^{i,m,\ell}, Y^{i,m,\ell-1})_i$ to ensure that the variance of the overall estimator is reduced. As for the iterative particle system, we require that $W^{i,m}$, $1 \le i \le N_{m,\ell}$, $m \in \mathbb{N}$, and $Y_0^{i,m,\ell}$, $1 \le i \le N_{m,\ell}$, $1 \le \ell \le L$, $m \in \mathbb{N}$, are independent.

3 Abstract framework for MLMC analysis

To streamline the analysis of the iterated MLMC estimator, we introduce an abstract framework corresponding to one iteration. This simplifies the notation and also may be useful for future developments of MLMC algorithms.

Let $\overline{b} : \mathbb{R}^d \times \mathcal{P}_2^s(\mathbb{R}^d) \to \mathbb{R}^d$ and $\overline{\sigma} : \mathbb{R}^d \times \mathcal{P}_2^s(\mathbb{R}^d) \to \mathbb{R}^{d\otimes r}$ be measurable functions. Also, $\mathcal{V} \in \mathcal{P}_2^s(C([0,T],\mathbb{R}^d))$ is fixed (the precise conditions that we impose on $\overline{b}$, $\overline{\sigma}$ and $\mathcal{V}$ will be presented in Section 3.1). We consider SDEs with *random* coefficients of the form

$$dU_t = \overline{b}(U_t, \mathcal{V}_t)dt + \overline{\sigma}(U_t, \mathcal{V}_t)dW_t, \quad \mu_0^U = \mu_0^X. \tag{3.1}$$

The solution of this SDE is well-defined under the assumptions in Section 3.1, by [22]. For $\ell = 1, \ldots, L$, the corresponding Euler approximation of (3.1) at level ℓ is given by

$$dZ_t^\ell = \overline{b}(Z_{\eta_\ell(t)}^\ell, \mathcal{V}_{\eta_\ell(t)})dt + \overline{\sigma}(Z_{\eta_\ell(t)}^\ell, \mathcal{V}_{\eta_\ell(t)})dW_t, \quad \mu_0^{Z^\ell} = \mu_0^X. \tag{3.2}$$

We require that $\mathcal{V}$ does not depend on ℓ and that $(W_t)_{t\in[0,T]}$ is independent of $\mathcal{V}$. Subsequently, we define a particle system $\{Z^{i,\ell}\}$ as follows,

$$dZ_t^{i,\ell} = \overline{b}(Z_{\eta_\ell(t)}^{i,\ell}, \mathcal{V}_{\eta_\ell(t)})dt + \overline{\sigma}(Z_{\eta_\ell(t)}^{i,\ell}, \mathcal{V}_{\eta_\ell(t)})dW_t^i, \qquad \mu_0^{Z^{i,\ell}} = \mu_0^X. \tag{3.3}$$

3.1 Analysis of the abstract framework

Using the notation defined in the previous section, we formulate the conditions needed to study the convergence of the iterated particle system. Recall that $\mathcal{V} \in \mathcal{P}_2^s(C([0,T],\mathbb{R}^d))$ is given and we consider equations (3.2) and (3.3). We assume the following.

Assumption 3.1.

($\mathcal{V}$-bound) *The random measure $\mathcal{V}$ is independent of W^i and $Z_0^{i,\ell}$. For each $p \geq 1$,*

$$\sup_{0\leq s\leq T} \mathbb{E}\left|\int_{\mathbb{R}^d} |y|^p \mathcal{V}_s(dy)\right| < \infty.$$

($\mathcal{V}$-Reg) *There exists a constant c such that*

$$\sup_{x\in\mathbb{R}^d} \sup_{0\leq s\leq t\leq T} \mathbb{E}\big[\big|\overline{b}(x,\mathcal{V}_t) - \overline{b}(x,\mathcal{V}_s)\big|^2 + \big\|\overline{\sigma}(x,\mathcal{V}_t) - \overline{\sigma}(x,\mathcal{V}_s)\big\|^2\big] \leq c(t-s).$$

($\mathcal{V}$-Lip) *There exists a constant c such that for each $t \in [0,T]$ and $x, y \in \mathbb{R}^d$,*

$$|\overline{b}(x,\mathcal{V}_t) - \overline{b}(y,\mathcal{V}_t)| + \|\overline{\sigma}(x,\mathcal{V}_t) - \overline{\sigma}(y,\mathcal{V}_t)\| \leq c|x-y| \tag{3.4}$$

$$|\overline{b}(x,\mathcal{V}_t)| + \|\overline{\sigma}(x,\mathcal{V}_t)\| \leq c\left(1 + |x| + \left|\int_{\mathbb{R}^d} |y|\,\mathcal{V}_t(dy)\right|\right). \tag{3.5}$$

Analysis of conditional MLMC variance For the rest of this section, we denote by c a generic constant that depends on T, but not on ℓ or N_ℓ. We first consider the integrability of process (3.2).

Lemma 3.2. *Let Z^ℓ be defined as in (3.2). Assume **($\mathcal{V}$-Lip)** and **(μ_0-L_p)** . Then for any $p \geq 2$ and $\ell \geq 0$, there exists a constant c such that*

$$\mathbb{E}\left[\sup_{t\in[0,T]} |Z_t^\ell|^p\right] \leq c\left(1 + \mathbb{E}\left[\int_0^T \left|\int_{\mathbb{R}^d} |y|^p \mathcal{V}_{\eta_\ell(s)}(dy)\right| ds\right]\right).$$

The proof is elementary and can be found in the Appendix A. The following two lemmas focus on the regularity of Z_t^ℓ in time and its strong convergence property. The first lemma bounds the difference in Z_t^ℓ over two time points, at a fixed level ℓ. The second lemma bounds the difference in Z_t^ℓ over adjacent levels, at a fixed time t. Their proofs follow from standard estimates in the theory of SDE and are therefore omitted.

Lemma 3.3 (Regularity of Z_t^ℓ)**.** *Let Z^ℓ be defined as in (3.2). Assume **($\mathcal{V}$-Lip)** and **($\mathcal{V}$-bound)** . Then, for $p \geq 1$, $0 \leq u \leq s \leq T$,*

$$\left(\mathbb{E}[|Z_s^\ell - Z_u^\ell|^p]\right)^{\frac{1}{p}} \leq c(s-u)^{\frac{1}{2}}.$$

Lemma 3.4 (Strong convergence of Z_t^ℓ)**.** *Assume* **($\mathcal{V}$-Lip), ($\mathcal{V}$-bound)** *and* **($\mathcal{V}$-Reg)** *. Then for any $\ell \in \{1, 2, \dots, L\}$, there exists a constant $c > 0$ such that*

$$\mathbb{E}\left[\sup_{0\le t\le T} |Z_t^\ell - Z_t^{\ell-1}|^2\right] \le ch_\ell.$$

We define the interpolated empirical measures $\widetilde{\mu}_t^{Z^\ell,N}$ exactly as in (2.3) and the corresponding MLMC operator $\mathcal{M}_t$ (corresponding to (2.4), but for one Picard iteration) as

$$\langle \mathcal{M}_t, P(x,\cdot)\rangle = \left\langle \sum_{\ell=0}^{L} \left(\widetilde{\mu}_t^{Z^\ell,N_\ell} - \widetilde{\mu}_t^{Z^{\ell-1},N_\ell}\right), P(x,\cdot)\right\rangle, \qquad \widetilde{\mu}_t^{Z^{-1},N_0} := 0.$$

We also define σ-algebra $\mathcal{F}_t^{\mathcal{V}} = \{\sigma(\mathcal{V}_s)_{0\le s\le t}\}$. Since samples $\{Z_{\eta_L(t)}^{i,\ell}\}$, $i = 1, \dots, N_\ell$, $\ell = 0, \dots, L$, conditioned on $\mathcal{F}_T^{\mathcal{V}}$ are independent, we can bound the conditional MLMC variance as follows.

Lemma 3.5. *Assume* **($\mathcal{V}$-Lip), ($\mathcal{V}$-bound)** *and* **($\mathcal{V}$-Reg)** *hold. Let $\mu \in \mathcal{P}_2\big(C([0,T],\mathbb{R}^d)\big)$. Then for any Lipschitz function $P : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, there exists a constant c such that*

$$\sup_{0\le t\le T} \int_{\mathbb{R}^d} \mathbb{E}\left[\mathrm{Var}\left(\langle \mathcal{M}_{\eta_L(t)}, P(x,\cdot)\rangle \middle| \mathcal{F}_T^{\mathcal{V}}\right)\right] \mu_t(dx) \le c \sum_{\ell=0}^{L} \frac{h_\ell}{N_\ell}. \tag{3.6}$$

Proof. The independence condition in **($\mathcal{V}$-bound)** implies that

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{Var}\left(\langle \mathcal{M}_{\eta_L(t)}, P(x,\cdot)\rangle \middle| \mathcal{F}_T^{\mathcal{V}}\right)\right] \\
&= \sum_{i=1}^{N_0} \frac{1}{N_0^2} \mathbb{E}\left[\mathrm{Var}\left[P_{\eta_L(t)}^{i,0} \middle| \mathcal{F}_T^{\mathcal{V}}\right]\right] + \sum_{\ell=1}^{L}\sum_{i=1}^{N_\ell} \frac{1}{N_\ell^2} \mathbb{E}\left[\mathrm{Var}\left[P_{\eta_L(t)}^{i,\ell} - P_{\eta_L(t)}^{i,\ell-1} \middle| \mathcal{F}_T^{\mathcal{V}}\right]\right],
\end{aligned}$$

where

$$P_{\eta_L(t)}^{i,\ell} := (1-\lambda_t^\ell) P(x, Z_{\eta_\ell(\eta_L(t))}^{i,\ell}) + \lambda_t^\ell P(x, Z_{\eta_\ell(\eta_L(t))+h_\ell}^{i,\ell}), \tag{3.7}$$

$\lambda_t^\ell = \frac{\eta_L(t) - \eta_\ell(\eta_L(t))}{h_\ell} \in [0,1]$. Using the fact that $\mathbb{E}[\mathrm{Var}(X|\mathcal{G})] \le \mathrm{Var}(X) \le \mathbb{E}[X^2]$, we obtain the bound

$$\mathbb{E}\left[\mathrm{Var}\left(\langle \mathcal{M}_{\eta_L(t)}, P(x,\cdot)\rangle \middle| \mathcal{F}_T^{\mathcal{V}}\right)\right] \le \sum_{i=1}^{N_0} \frac{1}{N_0^2} \mathbb{E}\left|P_{\eta_L(t)}^{i,0}\right|^2 + \sum_{\ell=1}^{L}\sum_{i=1}^{N_\ell} \frac{1}{N_\ell^2} \mathbb{E}\left|P_{\eta_L(t)}^{i,\ell} - P_{\eta_L(t)}^{i,\ell-1}\right|^2.$$

Since P is Lipschitz, it has linear growth. By Lemma 3.2, it follows that

$$\mathbb{E}\left|P_{\eta_L(t)}^{i,0}\right|^2 \le c \sup_{0\le t\le T} \int_{\mathbb{R}^d} \left(x^2 + \mathbb{E}|Z_{\eta_0(\eta_L(t))}^{i,0}|^2 + \mathbb{E}|Z_{\eta_0(\eta_L(t))+h_0}^{i,0}|^2\right) \mu_t(dx) < +\infty.$$

Next, we consider levels $\ell \in \{1, \dots, L\}$. Recall from (3.7) that

$$\begin{aligned} P^{i,\ell}_{\eta_L(t)} &= (1-\lambda^\ell_t) P(x, Z^{i,\ell}_{\eta_\ell(\eta_L(t))}) + \lambda^\ell_t P(x, Z^{i,\ell}_{\eta_\ell(\eta_L(t))+h_\ell}), \\ P^{i,\ell-1}_{\eta_L(t)} &= (1-\lambda^{\ell-1}_t) P(x, Z^{i,\ell-1}_{\eta_{\ell-1}(\eta_L(t))}) + \lambda^{\ell-1}_t P(x, Z^{i,\ell-1}_{\eta_{\ell-1}(\eta_L(t))+h_{\ell-1}}). \end{aligned}$$

We decompose the error as follows.

$$\begin{aligned} & |P^{i,\ell}_{\eta_L(t)} - P^{i,\ell-1}_{\eta_L(t)}| \\ \leq \;& (1-\lambda^{\ell-1}_t) \cdot \left| P(x, Z^{i,\ell}_{\eta_\ell(\eta_L(t))}) \pm P(x, Z^{i,\ell-1}_{\eta_\ell(\eta_L(t))}) - P(x, Z^{i,\ell-1}_{\eta_{\ell-1}(\eta_L(t))}) \right| \\ & + \lambda^{\ell-1}_t \cdot \left| P(x, Z^{i,\ell}_{\eta_\ell(\eta_L(t))+h_\ell}) \pm P(x, Z^{i,\ell-1}_{\eta_\ell(\eta_L(t))+h_\ell}) - P(x, Z^{i,\ell-1}_{\eta_{\ell-1}(\eta_L(t))+h_{\ell-1}}) \right| \\ & + |\lambda^\ell_t - \lambda^{\ell-1}_t| \cdot \left| P(x, Z^{i,\ell}_{\eta_\ell(\eta_L(t))+h_\ell}) - P(x, Z^{i,\ell}_{\eta_\ell(\eta_L(t))}) \right|. \end{aligned}$$

By Lemma 3.4,

$$\begin{aligned} \mathbb{E}|P(x, Z^{i,\ell}_{\eta_\ell(\eta_L(t))}) - P(x, Z^{i,\ell-1}_{\eta_\ell(\eta_L(t))})|^2 &\leq c h_\ell, && (3.8) \\ \mathbb{E}|P(x, Z^{i,\ell}_{\eta_\ell(\eta_L(t))+h_\ell}) - P(x, Z^{i,\ell-1}_{\eta_\ell(\eta_L(t))+h_\ell})|^2 &\leq c h_\ell. && (3.9) \end{aligned}$$

Also, by Lemma 3.3,

$$\mathbb{E}|P(x, Z^{i,\ell-1}_{\eta_\ell(\eta_L(t))}) - P(x, Z^{i,\ell-1}_{\eta_{\ell-1}(\eta_L(t))})|^2 \leq c(\eta_\ell(\eta_L(t)) - \eta_{\ell-1}(\eta_L(t))) \leq c h_\ell, \tag{3.10}$$

$$\mathbb{E}|P(x, Z^{i,\ell-1}_{\eta_\ell(\eta_L(t))+h_\ell}) - P(x, Z^{i,\ell-1}_{\eta_{\ell-1}(\eta_L(t))+h_{\ell-1}})|^2 \leq c h_\ell, \tag{3.11}$$

and

$$\mathbb{E}|P(x, Z^{i,\ell}_{\eta_\ell(\eta_L(t))+h_\ell}) - P(x, Z^{i,\ell}_{\eta_\ell(\eta_L(t))})|^2 \leq c h_\ell. \tag{3.12}$$

We obtain (3.6) by combining (3.8), (3.9), (3.10), (3.11) and (3.12). Since t and x are arbitrary, the proof is complete.

□

3.2 Weak error analysis

We begin this subsection by defining $\mathcal{X}^{s,x}$ as

$$\mathcal{X}^{s,x}_t = x + \int_s^t b[\mathcal{X}^{s,x}_u, \mu^X_u]\, du + \int_s^t \sigma[\mathcal{X}^{s,x}_u, \mu^X_u]\, dW_u.$$

For $P \in C^{0,2}_{b,b}(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R})$ and $t \in [0,T]$, we consider the function

$$v_y(s,x) := \mathbb{E}[P(y, \mathcal{X}^{s,x}_t)], \quad y \in \mathbb{R}^d \text{ and } (s,x) \in [0,t] \times \mathbb{R}^d. \tag{3.13}$$

We aim to show that $v_y(s,x) \in C^{1,2}$. The first step is the lemma below.

Lemma 3.6. *Assume* (μ_0-L_p) *and* (**Ker-Reg**) *. Then*

$$b[\cdot,\mu_\cdot^X]\in C^{2,1}_{b,b}(\mathbb{R}^d\times[0,T],\mathbb{R}^d) \text{ and } \sigma[\cdot,\mu_\cdot^X]\in C^{2,1}_{b,b}(\mathbb{R}^d\times[0,T],\mathbb{R}^{d\otimes r}).$$

Proof. For any $x\in\mathbb{R}^d$, $s\in[0,T]$ and $t\in[s,T]$, we apply Itô's formula to each coordinate $k\in\{1,\dots,d\}$ of b to get

$$\begin{aligned} b_k(x,X_t) &= b_k(x,X_s)+\int_s^t\sum_{j=1}^d\sum_{i=1}^r\partial_{y_j}b_k(x,X_u)\sigma_{ji}[X_u,\mu_u^X]dW_u^i \\ &\quad+\int_s^t\sum_{j=1}^d\partial_{y_j}b_k(x,X_u)b_j[X_u,\mu_u^X]du+\frac{1}{2}\int_s^t\sum_{i,j=1}^d\partial^2_{y_i,y_j}b_k(x,X_u)a_{ij}[X_u,\mu_u^X]du, \end{aligned} \tag{3.14}$$

where $a[x,\mu]=\sigma[x,\mu]\sigma[x,\mu]^T$ and $\partial_{y_i}b_k,\partial^2_{y_i,y_j}b_k$ indicate the derivatives w.r.t. the the second argument. Assumptions **(Ker-Reg)** , **(Lip)** , (μ_0-L_p) and (2.1) imply that the above stochastic integral is a martingale. By the fundamental theorem of calculus,

$$\partial_t\mathbb{E}[b_k(x,X_t)]=\mathbb{E}\Bigg[\sum_{j=1}^d\partial_{y_j}b_k(x,X_t)b_j[x,\mu_t^X]+\frac{1}{2}\sum_{i,j=1}^d\partial^2_{y_i,y_j}b_k(x,X_t)a_{ij}[x,\mu_t^X]\Bigg]. \tag{3.15}$$

By **(Ker-Reg)** , $\partial_{y_j}b_k$ and $\partial^2_{y_i,y_j}b_k$ are bounded. Moreover, by **(Lip)** , we know that b and a are respectively of linear and quadratic growth in x. Therefore, by (2.1), we conclude that $\partial_t b_k[x,\mu_t^X]$ is bounded. To conclude, we can apply the same argument to $\sigma[\cdot,\mu_\cdot^X]$. ☐

Lemma 3.7. *Assume* (**Ker-Reg**) *and* (μ_0-L_p) *. Then for any* $(s,x)\in[0,t]\times\mathbb{R}^d$*,* $(i,j)\in\{1,\dots,d\}^2$ *and* $P\in C^{0,2}_{b,b}(\mathbb{R}^d\times\mathbb{R}^d,\mathbb{R})$*,*

$$\sup_{y\in\mathbb{R}^d}(\|\partial_{x_i}v_y(s,x)\|_\infty+\|\partial^2_{x_i,x_j}v_y(s,x)\|_\infty\le L. \tag{(v-diff-Reg+)}$$

Proof. We only provide a sketch as the argument is standard. By the fact that the first-order spatial derivatives of $b[\cdot,\mu_\cdot^X]$ and $\sigma[\cdot,\mu_\cdot^X]$ are bounded, it is straightforward to deduce that

$$\sup_{x\in\mathbb{R}^d}\sup_{s\in[0,t]}\mathbb{E}\left[\left|\partial_{x_i}(X_t^{s,x})^{(j)}\right|^2\right]<\infty. \tag{3.16}$$

Theorem 5.5.5 in [15] establishes that

$$\partial_{x_i}v_y(s,x)=\sum_{j=1}^d\mathbb{E}\Big[\partial_{y_j}P(y,X_t^{s,x})\partial_{x_i}(X_t^{s,x})^{(j)}\Big]. \tag{3.17}$$

By (3.16), it is clear that the assertion for the first order derivatives in (**(v-diff-Reg+)**) holds if $P\in C^{0,2}_{b,b}(\mathbb{R}^d\times\mathbb{R}^d,\mathbb{R})$. Similarly, we can prove the assertion for the second order derivatives in the same way. ☐

By the Feynman-Kac theorem ([21]), it can be shown that $v_y(\cdot,\cdot)$ satisfies the following Cauchy problem,

$$\begin{cases} \partial_s v_y(s,x) + \frac{1}{2}\sum_{i,j=1}^{d}\Big(\sigma[x,\mu_s^X]\sigma[x,\mu_s^X]^T\Big)_{ij}\partial^2_{x_i,x_j}v_y(s,x) & \\ \qquad\qquad\qquad + \sum_{j=1}^{d}\Big(b[x,\mu_s^X]\Big)_j\partial_{x_j}v_y(s,x) &= 0, \quad (s,x)\in[0,t]\times\mathbb{R}^d, \\ \qquad\qquad\qquad\qquad\qquad\qquad v_y(t,x) &= P(y,x). \end{cases} \tag{3.18}$$

The following theorem reveals the order of weak convergence of (3.2) to (1.1). We denote by $\mu_t^{Z^\ell|\mathcal{F}_T^{\mathcal{V}}}$ the regular conditional probability measure of Z_t^ℓ given $F_T^{\mathcal{V}}$. (See Theorem 7.1 in [27] for details.) The existence of regular conditional probability measure follows from the fact that we work on a Polish space with the Borel $\sigma-$algebra.

Theorem 3.8. *Let $P \in C_{b,b}^{0,2}(\mathbb{R}^d\times\mathbb{R}^d,\mathbb{R})$ be a Lipschitz continuous function.* [1] *Assume that **(Ker-Reg)** , **(μ_0-L_p)** , **($\mathcal{V}$-bound)** and **($\mathcal{V}$-Lip)** hold. Then there exists a constant c (independent of the choices of L and $N_1,\dots,N_L$) such that for each $t\in[0,T]$, $\ell\in\{0,\dots,L\}$ and $x\in\mathbb{R}^d$,*

$$\begin{aligned} &\sup_{0\le s\le t}|\mathbb{E}[P(x,Z_s^\ell)]-\mathbb{E}[P(x,X_s)]| \\ &\quad\le c\bigg(h_\ell + \int_0^t \mathbb{E}\Big[\int_{\mathbb{R}^d}\Big|\overline{b}(x,\mathcal{V}_{\eta_\ell(s)})-\mathbb{E}[b(x,X_{\eta_\ell(s)})]\Big|\,\mu_{\eta_\ell(s)}^{Z^\ell|\mathcal{F}_T^{\mathcal{V}}}(dx)\Big]ds \\ &\qquad + \int_0^t \mathbb{E}\Big[\int_{\mathbb{R}^d}\Big\|\overline{\sigma}(x,\mathcal{V}_{\eta_\ell(s)})-\mathbb{E}[\sigma(x,X_{\eta_\ell(s)})]\Big\|\,\mu_{\eta_\ell(s)}^{Z^\ell|\mathcal{F}_T^{\mathcal{V}}}(dx)\Big]ds\bigg). \end{aligned}$$

Proof. To lighten the notation in this proof, we use t_k, $\eta(s)$ and Z to denote t_k^ℓ, $\eta_\ell(s)$ and Z^ℓ respectively. First, we observe that

$$|\mathbb{E}[P(y,Z_s)]-\mathbb{E}[P(y,X_s)]| \le \mathbb{E}|\mathbb{E}[P(y,Z_s)|\mathcal{F}_T^{\mathcal{V}}]-\mathbb{E}[P(y,X_s)]|.$$

From definition of $v(\cdot,\cdot)$ in (3.13), we compute that

$$\begin{aligned} \mathbb{E}[v_y(0,X_0)] = \int_{\mathbb{R}^d} v_y(0,x)\,\mu_0(dx) &= \int_{\mathbb{R}^d}\mathbb{E}\big[P(y,\mathcal{X}_t^{0,x})\big]\,\mu_0(dx) \\ &= \mathbb{E}\big[\mathbb{E}[P(y,X_t)|X_0]\big]. \end{aligned}$$

[1] Note that the regularity of P can be relaxed to $C_{b,p}^{0,2}(\mathbb{R}^d\times\mathbb{R}^d,\mathbb{R})$. We prove the result in a slightly stronger assumption for the sake of simplicity.

The Feynman-Kac theorem, hypothesis **($\mathcal{V}$-bound)** and the fact that $\mu_0^X = \mu_0^Z$ give

$$\begin{aligned}\mathbb{E}[P(y,Z_t)|\mathcal{F}_T^{\mathcal{V}}] - \mathbb{E}[P(y,X_t)] &= \mathbb{E}[v_y(t,Z_t)|\mathcal{F}_T^{\mathcal{V}}] - \mathbb{E}[v_y(0,Z_0)] \\ &= \mathbb{E}[v_y(t,Z_t)|\mathcal{F}_T^{\mathcal{V}}] - \mathbb{E}[v_y(0,Z_0)|\mathcal{F}_T^{\mathcal{V}}] \\ &= \sum_{k=0}^{n-1} \mathbb{E}\big[v_y(t_{k+1},Z_{k+1}) - v_y(t_k,Z_k)\,\big|\,\mathcal{F}_T^{\mathcal{V}}\big],\end{aligned}$$

where $n = t/h_\ell$ [2]. By Itô's formula,

$$\begin{aligned}&\mathbb{E}[v_y(t,Z_t)|\mathcal{F}_T^{\mathcal{V}}] - \mathbb{E}[v_y(0,Z_0)] \\ = &\sum_{k=0}^{n-1}\mathbb{E}\Bigg[\int_{t_k}^{t_{k+1}}\Bigg(\partial_t v_y(s,Z_s) + \sum_{j=1}^d \partial_{x_j} v_y(s,Z_s)\overline{b}_j(Z_{\eta(s)},\mathcal{V}_{\eta(s)}) \\ &+\frac{1}{2}\sum_{i,j=1}^d \partial^2_{x_i,x_j} v_y(s,Z_s)\overline{a}_{ij}(Z_{\eta(s)},\mathcal{V}_{\eta(s)})\Bigg)ds \\ &+\int_{t_k}^{t_{k+1}}\sum_{j=1}^d\sum_{i=1}^r \partial_{x_j} v_y(s,Z_s)\overline{\sigma}_{ji}(Z_{\eta(s)},\mathcal{V}_{\eta(s)})dW_s^{(i)}\,\Bigg|\,\mathcal{F}_T^{\mathcal{V}}\Bigg],\end{aligned}$$

where $\overline{a}(x,\mu) = \overline{\sigma}(x,\mu)\overline{\sigma}(x,\mu)^T$. Condition **((v-diff-Reg+))**, as well as hypotheses **(Lip)** , **(μ_0-L_p)** and **($\mathcal{V}$-bound)** , along with Lemma 3.2 and part (a) of Lemma A.1 (with the filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ such that $\mathcal{F}_t = \sigma(\mathcal{F}_T^{\mathcal{V}}, \{W_u\}_{0\le u\le t}, \{Z_u\}_{0\le u\le t})$) imply that

$$\mathbb{E}\Bigg[\int_{t_k}^{t_{k+1}}\sum_{j=1}^d\sum_{i=1}^r \partial_{x_j} v_y(s,Z_s)\overline{\sigma}_{ji}(Z_{\eta(s)},\mathcal{V}_{\eta(s)})dW_s^{(i)}\,\Bigg|\,\mathcal{F}_T^{\mathcal{V}}\Bigg] = 0. \tag{3.19}$$

Subsequently, using the fact that $v(\cdot,\cdot)$ satisfies PDE (3.18), we have

$$\begin{aligned}&\mathbb{E}[v_y(t,Z_t)|\mathcal{F}_T^{\mathcal{V}}] - \mathbb{E}[v_y(0,Z_0)] \\ = &\sum_{k=0}^{n-1}\int_{t_k}^{t_{k+1}}\mathbb{E}\Bigg[\sum_{j=1}^d \partial_{x_j} v_y(s,Z_s)(\overline{b}_j(Z_{\eta(s)},\mathcal{V}_{\eta(s)}) - b_j[Z_s,\mu_s^X]) \\ &+\frac{1}{2}\sum_{i,j=1}^d \partial^2_{x_i,x_j} v_y(s,Z_s)(\overline{a}_{ij}(Z_{\eta(s)},\mathcal{V}_{\eta(s)}) - a_{ij}[Z_s,\mu_s^X])\,\Bigg|\,\mathcal{F}_T^{\mathcal{V}}\Bigg]ds.\end{aligned}$$

Hence,

$$\mathbb{E}[v_y(t,Z_t)|\mathcal{F}_T^{\mathcal{V}}] - \mathbb{E}[v_y(0,Z_0)] = \sum_{k=0}^{n-1}\int_{t_k}^{t_{k+1}}\mathbb{E}\Bigg[\sum_{i=1}^4 R_i(s)\Bigg|\mathcal{F}_T^{\mathcal{V}}\Bigg]ds,$$

[2] For simplicity we assume that n is an integer.

where

$$R_1(s) := \sum_{j=1}^{d} \partial_{x_j} v_y(s, Z_s)(b_j[Z_{\eta(s)}, \mu^X_{\eta(s)}] - b_j[Z_s, \mu^X_s])$$

$$R_2(s) := \sum_{j=1}^{d} \partial_{x_j} v_y(s, Z_s)(\overline{b}_j(Z_{\eta(s)}, \mathcal{V}_{\eta(s)}) - b_j[Z_{\eta(s)}, \mu^X_{\eta(s)}])$$

$$R_3(s) := \frac{1}{2} \sum_{i,j=1}^{d} \partial^2_{x_i,x_j} v_y(s, Z_s)(a_{ij}[Z_{\eta(s)}, \mu^X_{\eta(s)}] - a_{ij}[Z_s, \mu^X_s])$$

$$R_4(s) := \frac{1}{2} \sum_{i,j=1}^{d} \partial^2_{x_i,x_j} v_y(s, Z_s)(\overline{a}_{ij}(Z_{\eta(s)}, \mathcal{V}_{\eta(s)}) - a_{ij}[Z_{\eta(s)}, \mu^X_{\eta(s)}]).$$

Error R_1: Let $\mathcal{F}^Z_T$ be the sigma-algebra generated by $\{Z_t\}_{t\in[0,T]}$. From part (a) of Lemma A.1 and the Itô's formula, we have

$$\begin{aligned}
&\mathbb{E}[R_1(s)|\mathcal{F}^{\mathcal{V}}_T] \\
= &\sum_{k=1}^{d} \mathbb{E}\Big[\partial_{x_k} v_y(s, Z_s)\mathbb{E}\Big[\int_{\eta(s)}^{s} \Big[\partial_u b_k[Z_u, \mu^X_u] + \sum_{i=1}^{d} \partial_{x_i} b_k[Z_u, \mu^X_u]\overline{b}_i(Z_{\eta(u)}, \mathcal{V}_{\eta(u)}) + \\
&+\frac{1}{2} \sum_{i,j=1}^{d} \partial^2_{x_i,x_j} b_k[Z_u, \mu^X_u]\overline{a}_{ij}(Z_{\eta(u)}, \mathcal{V}_{\eta(u)})\Big]\, du \,\Big|\sigma(\mathcal{F}^Z_T, \mathcal{F}^{\mathcal{V}}_T)\Big] \,\Big|\mathcal{F}^{\mathcal{V}}_T\Big].
\end{aligned}$$

Condition ((**v-diff-Reg+**)) and the conditional Jensen inequality imply that

$$\begin{aligned}
&\mathbb{E}\big|\mathbb{E}[R_1(s)|\mathcal{F}^{\mathcal{V}}_T]\big| \\
\leq\ & c\sum_{k=1}^{d}\Big(\int_{\eta(s)}^{s} \mathbb{E}\Big|\partial_u b_k[Z_u, \mu^X_u] + \sum_{i=1}^{d} \partial_{x_i} b_k[Z_u, \mu^X_u]\overline{b}_i(Z_{\eta(u)}, \mathcal{V}_{\eta(u)}) + \\
&\frac{1}{2} \sum_{i,j=1}^{d} \partial^2_{x_i,x_j} b_k[Z_u, \mu^X_u]\overline{a}_{ij}(Z_{\eta(u)}, \mathcal{V}_{\eta(u)})\Big|du\Big).
\end{aligned} \tag{3.20}$$

Using these two bounds along with Lemma 3.6 and assumption (**$\mathcal{V}$-Lip**), we can see that

$$\mathbb{E}\big|\mathbb{E}[R_1(s)|\mathcal{F}^{\mathcal{V}}_T]\big| \leq c\Big(\int_{\eta(s)}^{s} 1 + \sup_{s'\in[0,t]} \mathbb{E}|Z_{s'}|^2 + \sup_{s'\in[0,t]} \mathbb{E}\Big|\int_{\mathbb{R}^d} |x|^2 \mathcal{V}_{s'}(dx)\Big|\, du\Big).$$

Assumptions **(Lip)** , (μ_0-L_p) and (**$\mathcal{V}$-bound**) allow us to conclude that

$$\sup_{0\leq s\leq t} \mathbb{E}\big|\mathbb{E}[R_1(s)|\mathcal{F}^{\mathcal{V}}_T]\big| \leq ch_\ell.$$

Error R_2: Condition ((**v-diff-Reg+**)) implies that

$$\left|\mathbb{E}[R_2(s)|\mathcal{F}_T^{\mathcal{V}}]\right| \leq c\,\mathbb{E}\big[|b[Z_{\eta(s)},\mu^X_{\eta(s)}] - \overline{b}(Z_{\eta(s)},\mathcal{V}_{\eta(s)})|\,\big|\mathcal{F}_T^{\mathcal{V}}\big].$$

Using the notation of regular conditional probability measures,

$$\mathbb{E}|\mathbb{E}[R_2(s)|\mathcal{F}_T^{\mathcal{V}}]| \leq c\,\mathbb{E}\bigg[\int_{\mathbb{R}^d}\left|\mathbb{E}[b(x,X_{\eta(s)})] - \overline{b}(x,\mathcal{V}_{\eta(s)})\right|\mu^{Z|\mathcal{F}_T^{\mathcal{V}}}_{\eta(s)}(dx)\bigg].$$

Similarly, by the condition on the second-order derivatives from ((**v-diff-Reg+**)), we can establish that

$$\sup_{0\leq s\leq T}\mathbb{E}|\mathbb{E}[R_3(s)|\mathcal{F}_T^{\mathcal{V}}]| \leq ch_\ell \tag{3.21}$$

and

$$|\mathbb{E}[R_4(s)|\mathcal{F}_T^{\mathcal{V}}]| \leq c\,\mathbb{E}\big[\|\sigma[Z_{\eta(s)},\mu^X_{\eta(s)}] - \overline{\sigma}(Z_{\eta(s)},\mathcal{V}_{\eta(s)})\|\,\big|\mathcal{F}_T^{\mathcal{V}}\big]. \tag{3.22}$$

□

Next, we introduce an artificial process $\bar{Z}^\ell$ in order to remove the dependence of Z^ℓ on $\mathcal{F}_T^{\mathcal{V}}$. Note that $\mu^{Z^\ell|\mathcal{F}_T^{\mathcal{V}}}_{\eta_\ell(s)}$ is a random measure, whereas $\mu^{\bar{Z}^\ell}_{\eta_\ell(s)}$ is non-random. This is crucial in the iteration that will be discussed in the next section.

Lemma 3.9. *Let $P \in C^{0,2}_{b,b}(\mathbb{R}^d\times\mathbb{R}^d,\mathbb{R})$ be a Lipschitz continuous function. Assume that* ***(Ker-Reg)*** *,* ***(μ_0-L_p)*** *,* ***($\mathcal{V}$-bound)*** *and* ***($\mathcal{V}$-Lip)*** *hold. Then there exists a constant c (independent of the choices of L and $N_1,\dots,N_L$) such that for each $t\in[0,T]$, $\ell\in\{0,\dots,L\}$ and $x\in\mathbb{R}^d$,*

$$\begin{aligned}
&\sup_{0\leq s\leq t}\mathbb{E}\Big[|\mathbb{E}[P(x,Z^\ell_s)|\mathcal{F}_T^{\mathcal{V}}] - \mathbb{E}[P(x,X_s)]|^2\Big] \\
\leq\ & c\bigg(h_\ell^2 + \int_0^t\bigg[\int_{\mathbb{R}^d}\mathbb{E}|\overline{b}(x,\mathcal{V}_{\eta_\ell(s)}) - \mathbb{E}[b(x,X_{\eta_\ell(s)})]|^2\mu^{\bar{Z}^\ell}_{\eta_\ell(s)}(dx)\bigg]ds \\
&+\int_0^t\bigg[\int_{\mathbb{R}^d}\mathbb{E}\Big\|\overline{\sigma}(x,\mathcal{V}_{\eta_\ell(s)}) - \mathbb{E}[\sigma(x,X_{\eta_\ell(s)})]\Big\|^2\mu^{\bar{Z}^\ell}_{\eta_\ell(s)}(dx)\bigg]ds\bigg),
\end{aligned}$$

where $\bar{Z}^\ell$ is a process defined by

$$d\bar{Z}^\ell_t = \int_{\mathbb{R}^d} b(\bar{Z}^\ell_{\eta_\ell(t)},y)\,\mu^X_{\eta_\ell(t)}(dy)\,dt + \int_{\mathbb{R}^d}\sigma(\bar{Z}^\ell_{\eta_\ell(t)},y)\,\mu^X_{\eta_\ell(t)}(dy)\,dW_t.$$

Proof. As in the proof of Theorem 3.8, we use $\eta(s)$, Z and $\bar{Z}$ to denote $\eta_\ell(s)$, Z^ℓ and $\bar{Z}^\ell$ respectively. By **(Lip)** and **($\mathcal{V}$-Lip)**,

$$\begin{aligned}
&\mathbb{E}\Big[\Big|\big(b[Z_{\eta(s)},\mu^X_{\eta(s)}] - \overline{b}(Z_{\eta(s)},\mathcal{V}_{\eta(s)})\big) - \big(b[\bar{Z}_{\eta(s)},\mu^X_{\eta(s)}] - \overline{b}(\bar{Z}_{\eta(s)},\mathcal{V}_{\eta(s)})\big)\Big|^2\,\Big|\mathcal{F}_T^{\mathcal{V}}\Big] \\
&\leq c\,\mathbb{E}\big[|Z_{\eta(s)} - \bar{Z}_{\eta(s)}|^2\,\big|\mathcal{F}_T^{\mathcal{V}}\big].
\end{aligned} \tag{3.23}$$

We further decompose the error as follows.

$$
\begin{aligned}
\mathbb{E}\big[\big|Z_{\eta(s)}-\bar{Z}_{\eta(s)}\big|^2\,\big|\mathcal{F}_T^{\mathcal{V}}\big] \quad &\leq \quad 2\Bigg(\mathbb{E}\Bigg[\bigg|\int_0^s\Big(b[\bar{Z}_{\eta(u)},\mu_{\eta(u)}^X]-\bar{b}\big(Z_{\eta(u)},\mathcal{V}_{\eta(u)}\big)\Big)\,du\bigg|^2\,\bigg|\mathcal{F}_T^{\mathcal{V}}\Bigg] \\
&\quad +\mathbb{E}\Bigg[\bigg|\int_0^s\Big(\sigma[\bar{Z}_{\eta(u)},\mu_{\eta(u)}^X]-\overline{\sigma}\big(Z_{\eta(u)},\mathcal{V}_{\eta(u)}\big)\Big)\,dW_u\bigg|^2\,\bigg|\mathcal{F}_T^{\mathcal{V}}\Bigg]\Bigg) \\
&=: \quad 2(R_{21}(s)+R_{22}(s)).
\end{aligned}
$$

By the conditional Fubini's theorem and the Cauchy-Schwarz inequality, there exists a constant $K>0$ such that

$$
\begin{aligned}
& R_{21}(s) \\
\leq\ & c\Bigg(\int_0^s\mathbb{E}\bigg[\Big|b[\bar{Z}_{\eta(u)},\mu_{\eta(u)}^X]-\bar{b}\big(\bar{Z}_{\eta(u)},\mathcal{V}_{\eta(u)}\big)\Big|^2\Big|\mathcal{F}_T^{\mathcal{V}}\bigg] \\
& +\mathbb{E}\bigg[\Big|\bar{b}\big(\bar{Z}_{\eta(u)},\mathcal{V}_{\eta(u)}\big)-\bar{b}\big(Z_{\eta(u)},\mathcal{V}_{\eta(u)}\big)\Big|^2\Big|\mathcal{F}_T^{\mathcal{V}}\bigg]\,du\Bigg) \\
\leq\ & c\Bigg(\int_0^s\mathbb{E}\bigg[\Big|b[\bar{Z}_{\eta(u)},\mu_{\eta(u)}^X]-\bar{b}\big(\bar{Z}_{\eta(u)},\mathcal{V}_{\eta(u)}\big)\Big|^2\Big|\mathcal{F}_T^{\mathcal{V}}\bigg]+\mathbb{E}\big[\big|Z_{\eta(u)}-\bar{Z}_{\eta(u)}\big|^2\big|\mathcal{F}_T^{\mathcal{V}}\big]\,du\Bigg),
\end{aligned}
$$

where assumption **($\mathcal{V}$-Lip)** is used in the final inequality. Since $\bar{Z}$ is independent of $\mathcal{F}_T^{\mathcal{V}}$ and that $\mu_{\eta(u)}^X$ is a non-random measure, we use the properties of regular conditional distributions as outlined in Theorem 7.1 of [27] to prove that for each $\omega\in\Omega$,

$$
\begin{aligned}
& \Bigg(\mathbb{E}\bigg[\Big|b[\bar{Z}_{\eta(u)},\mu_{\eta(u)}^X]-\bar{b}\big(\bar{Z}_{\eta(u)},\mathcal{V}_{\eta(u)}\big)\Big|^2\Big|\mathcal{F}_T^{\mathcal{V}}\bigg]\Bigg)(\omega) \\
=\ & \int_{\mathbb{R}^d}\Big|b[x,\mu_{\eta(u)}^X]-\bar{b}(x,\mathcal{V}_{\eta(u)}(\omega))\Big|^2\,\mu_{\eta(u)}^{\bar{Z}}(dx).
\end{aligned}
$$

Therefore,

$$
R_{21}(s)\leq c\bigg(\int_0^s\Big[\mathbb{E}\big[\big|Z_{\eta(u)}-\bar{Z}_{\eta(u)}\big|^2\big|\mathcal{F}_T^{\mathcal{V}}\big]+\int_{\mathbb{R}^d}\Big|b[x,\mu_{\eta(u)}^X]-\bar{b}(x,\mathcal{V}_{\eta(u)})\Big|^2\,\mu_{\eta(u)}^{\bar{Z}}(dx)\Big]\,du\bigg).
$$

We proceed similarly as $R_{22}(s)$ and apply part (b) of Lemma A.1 (with the filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ such that $\mathcal{F}_t=\sigma(\mathcal{F}_T^{\mathcal{V}},\{W_u\}_{0\leq u\leq t},Z_0)$) to get

$$
R_{22}(s)\leq c\bigg(\int_0^s\Big[\mathbb{E}\big[\big|Z_{\eta(u)}-\bar{Z}_{\eta(u)}\big|^2\big|\mathcal{F}_T^{\mathcal{V}}\big]+\int_{\mathbb{R}^d}\Big\|\sigma[x,\mu_{\eta(u)}^X]-\overline{\sigma}(x,\mathcal{V}_{\eta(u)})\Big\|^2\,\mu_{\eta(u)}^{\bar{Z}}(dx)\Big]\,du\bigg).
$$

Combining both bounds gives

$$
\begin{aligned}
\mathbb{E}\big[\big|Z_{\eta(s)}-\bar{Z}_{\eta(s)}\big|^2\big|\mathcal{F}_T^{\mathcal{V}}\big] \quad &\leq \quad c\bigg(\int_0^s\Big[\mathbb{E}\big[\big|Z_{\eta(u)}-\bar{Z}_{\eta(u)}\big|^2\big|\mathcal{F}_T^{\mathcal{V}}\big] \\
&\quad +\int_{\mathbb{R}^d}\Big|b[x,\mu_{\eta(u)}^X]-\bar{b}(x,\mathcal{V}_{\eta(u)})\Big|^2\,\mu_{\eta(u)}^{\bar{Z}}(dx) \\
&\quad +\int_{\mathbb{R}^d}\Big\|\sigma[x,\mu_{\eta(u)}^X]-\overline{\sigma}(x,\mathcal{V}_{\eta(u)})\Big\|^2\,\mu_{\eta(u)}^{\bar{Z}}(dx)\Big]\,du\bigg),
\end{aligned}
$$

for any $s \in [0,t]$. By Gronwall's lemma and integration from 0 to t in time, we obtain that

$$\begin{aligned}\int_0^t \mathbb{E}\big[\big|Z_{\eta(s)} - \bar{Z}_{\eta(s)}\big|^2 \big|\mathcal{F}_T^{\mathcal{V}}\big]\,ds \quad &\leq \quad c\Big(\int_0^t \Big[\int_{\mathbb{R}^d}\Big|b[x,\mu_{\eta(s)}^X] - \bar{b}(x,\mathcal{V}_{\eta(s)})\Big|^2 \mu_{\eta(s)}^{\bar{Z}}(dx) \\ &\quad + \int_{\mathbb{R}^d}\Big\|\sigma[x,\mu_{\eta(s)}^X] - \overline{\sigma}(x,\mathcal{V}_{\eta(s)})\Big\|^2 \mu_{\eta(s)}^{\bar{Z}}(dx)\Big]\,ds\Big).\end{aligned}$$

By (3.2) and (3.23), it is clear that

$$\begin{aligned}\int_0^t |\mathbb{E}[R_2(s)|\mathcal{F}_T^{\mathcal{V}}]|^2\,ds \quad &\leq \quad c\Big(\int_0^t \mathbb{E}\big[\big|Z_{\eta(s)} - \bar{Z}_{\eta(s)}\big|^2\big|\mathcal{F}_T^{\mathcal{V}}\big] \\ &\quad + \mathbb{E}\big[|b[\bar{Z}_{\eta(s)},\mu_{\eta(s)}^X] - \bar{b}(\bar{Z}_{\eta(s)},\mathcal{V}_{\eta(s)})|^2\,|\mathcal{F}_T^{\mathcal{V}}\big]\,ds\Big).\end{aligned}$$

This shows that

$$\begin{aligned}\int_0^t |\mathbb{E}[R_2(s)|\mathcal{F}_T^{\mathcal{V}}]|^2\,ds \quad &\leq \quad c\Big(\int_0^t \Big[\int_{\mathbb{R}^d}\Big|b[x,\mu_{\eta(s)}^X] - \bar{b}(x,\mathcal{V}_{\eta(s)})\Big|^2 \mu_{\eta(s)}^{\bar{Z}}(dx) \\ &\quad + \int_{\mathbb{R}^d}\Big\|\sigma[x,\mu_{\eta(s)}^X] - \overline{\sigma}(x,\mathcal{V}_{\eta(s)})\Big\|^2 \mu_{\eta(s)}^{\bar{Z}}(dx)\Big]\,ds\Big).\end{aligned}$$

We repeat the same argument for $R_4(s)$ and conclude that

$$\begin{aligned}\int_0^t |\mathbb{E}[R_4(s)|\mathcal{F}_T^{\mathcal{V}}]|^2\,ds \quad &\leq \quad c\Big(\int_0^t \Big[\int_{\mathbb{R}^d}\Big|b[x,\mu_{\eta(s)}^X] - \bar{b}(x,\mathcal{V}_{\eta(s)})\Big|^2 \mu_{\eta(s)}^{\bar{Z}}(dx) \\ &\quad + \int_{\mathbb{R}^d}\Big\|\sigma[x,\mu_{\eta(s)}^X] - \overline{\sigma}(x,\mathcal{V}_{\eta(s)})\Big\|^2 \mu_{\eta(s)}^{\bar{Z}}(dx)\Big]\,ds\Big).\end{aligned}$$

□

4 Iteration of the MLMC algorithm

4.1 Interacting kernels

Fix $m \geq 1$ and correspond each particle $Z^{i,\ell}$ in the abstract framework with $Y^{i,m,\ell}$ defined in (1.9) and $\mathcal{F}_T^{\mathcal{V}}$ with the sigma-algebra $\mathcal{F}^{m-1}$ generated by all the particles $Y^{i,m-1,\ell}$ in the $(m-1)$th Picard step, $0 \leq \ell \leq L, 1 \leq i \leq N_{m-1,\ell}$. We set $\mathcal{V}_t := \mathcal{M}_t^{(m-1)}$ (defined in (2.4)), $\overline{b}(x,\mu) := b[x,\mu]$ and $\overline{\sigma}(x,\mu) := \sigma[x,\mu]$, so that

$$\overline{b}(x,\mathcal{M}_t^{(m-1)}) = \langle \mathcal{M}_t^{(m-1)}, b(x,\cdot)\rangle \quad \text{and} \quad \overline{\sigma}(x,\mathcal{M}_t^{(m-1)}) = \langle \mathcal{M}_t^{(m-1)}, \sigma(x,\cdot)\rangle,$$

for each $x \in \mathbb{R}^d$. The measure $\mathcal{M}^{(m-1)}$ satisfies the independence criterion in **($\mathcal{V}$-bound)** , since $\{Y^{m-1}\} \perp (W^m, Z_0^m)$. The criteria **($\mathcal{V}$-bound)** , **($\mathcal{V}$-Reg)** and **($\mathcal{V}$-Lip)** are verified below.

In the results of this section, c denotes a generic constant that depends on T, but not on m,ℓ or $N_{m,\ell}$.

Lemma 4.1 (Verification of **($\mathcal{V}$-Lip)**)**.** *Assume **(Lip)** and **(μ_0-L_p)**. Then, for each $t \in [0,T]$, there exists a constant c such that for all $x_1, x_2 \in \mathbb{R}^d$*

$$|\langle \mathcal{M}_t^{(m-1)}, b(x_1,\cdot) - b(x_2,\cdot)\rangle| + \|\langle \mathcal{M}_t^{(m-1)}, \sigma(x_1,\cdot) - \sigma(x_2,\cdot)\rangle\| \le c|x_1 - x_2|,$$

$$|\langle \mathcal{M}_t^{(m-1)} b(x_1,\cdot)\rangle| + \|\langle \mathcal{M}_t^{(m-1)} \sigma(x_1,\cdot)\rangle\| \le c\Big(1 + |x| + \Big|\int_{\mathbb{R}^d} |y| \mathcal{M}_t^{(m-1)}(dy)\Big|\Big).$$

Proof. For any $t \in [0,T]$ and $x_1, x_2 \in \mathbb{R}^d$, by the definition of $\mathcal{M}_t^{(m-1)}$,

$$\begin{aligned}
&\Big|\langle \mathcal{M}_t^{(m-1)}, b(x_1,\cdot)\rangle - \langle \mathcal{M}_t^{(m-1)}, b(x_2,\cdot)\rangle\Big| \\
= \ &\Bigg|\sum_{\ell=1}^{L} \frac{1}{N_{m-1,\ell}} \sum_{i=1}^{N_{m-1,\ell}} \Bigg[\Big(\frac{t-\eta_\ell(t)}{h_\ell}\Big)\cdot\Big(b(x_1, Y^{i,m-1,\ell}_{\eta_\ell(t)+h_\ell}) - b(x_2, Y^{i,m-1,\ell}_{\eta_\ell(t)+h_\ell})\Big) \\
&+\Big(1-\frac{t-\eta_\ell(t)}{h_\ell}\Big)\cdot\Big(b(x_1, Y^{i,m-1,\ell}_{\eta_\ell(t)}) - b(x_2, Y^{i,m-1,\ell}_{\eta_\ell(t)})\Big) \\
&-\Big(\frac{t-\eta_{\ell-1}(t)}{h_{\ell-1}}\Big)\cdot\Big(b(x_1, Y^{i,m-1,\ell-1}_{\eta_{\ell-1}(t)+h_{\ell-1}}) - b(x_2, Y^{i,m-1,\ell-1}_{\eta_{\ell-1}(t)+h_{\ell-1}})\Big) \\
&-\Big(1-\frac{t-\eta_{\ell-1}(t)}{h_{\ell-1}}\Big)\cdot\Big(b(x_1, Y^{i,m-1,\ell-1}_{\eta_{\ell-1}(t)}) - b(x_2, Y^{i,m-1,\ell-1}_{\eta_{\ell-1}(t)})\Big)\Bigg] \\
&+\frac{1}{N_{m-1,0}} \sum_{i=1}^{N_{m-1,0}} \Bigg[\Big(\frac{t-\eta_0(t)}{h_0}\Big)\cdot\Big(b(x_1, Y^{i,m-1,0}_{\eta_0(t)+h_0}) - b(x_2, Y^{i,m-1,0}_{\eta_0(t)+h_0})\Big) \\
&+\Big(1-\frac{t-\eta_0(t)}{h_0}\Big)\cdot\Big(b(x_1, Y^{i,m-1,0}_{\eta_0(t)}) - b(x_2, Y^{i,m-1,0}_{\eta_0(t)})\Big)\Bigg]\Bigg|.
\end{aligned}$$

The required bounds follow from **(Lip)**. The corresponding estimates for $\|\overline{\sigma}(x_1, \mathcal{V}_{\eta(t)}) - \overline{\sigma}(x_2, \mathcal{V}_{\eta(t)})\|$ and $\|\overline{\sigma}(x_1, \mathcal{V}_{\eta(t)})\|$ can be obtained in a similar way and are hence omitted. ☐

Lemma 4.2 (Verification of **($\mathcal{V}$-bound)**)**.** *Assume **(Lip)** and **(μ_0-L_p)**. Then for any $p \ge 2$, there exists a constant c such that*

$$\sup_{n\in\mathbb{N}\cup\{0\}} \sup_{t\in[0,T]} \mathbb{E}\Big|\int_{\mathbb{R}^d} |x|^p \mathcal{M}_t^{(n)}(dx)\Big| \le c.$$

Proof. For simplicity of notation, we rewrite

$$\int_{\mathbb{R}^d} |x|^p \mathcal{M}_t^{(n)}(dx) := \frac{1}{N_0}\sum_{i=1}^{N_0} P_t^{i,0} + \sum_{\ell=1}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \Big(P_t^{i,\ell} - P_t^{i,\ell-1}\Big),$$

where

$$P_t^{i,\ell} = \Big(\frac{t-\eta_\ell(t)}{h_\ell}\Big)|Y^{i,n,\ell}_{\eta_\ell(t)+h_\ell}|^p + \Big(1-\frac{t-\eta_\ell(t)}{h_\ell}\Big)|Y^{i,n,\ell}_{\eta_\ell(t)}|^p.$$

We first fix $\ell \in \{1, \ldots, L\}$ and define

$$\Delta_t^{i,\ell} := \mathbb{E}|P_t^{i,\ell} - P_t^{i,\ell-1}|, \quad i \in \{1, \ldots, N_\ell\}.$$

By exchangeability, there exists a constant c (independent of the Picard step n) such that

$$\mathbb{E}[|\Delta_t^{i,\ell}|] \leq c \sum_{\ell'=\ell-1}^{\ell} (\mathbb{E}|Y_{\eta_{\ell'}(t)}^{1,n,\ell'}|^p + \mathbb{E}|Y_{\eta_{\ell'}(t)+h_{\ell'}}^{1,n,\ell'}|^p).$$

By the triangle inequality,

$$\mathbb{E}\left|\frac{1}{N_\ell}\sum_{i=1}^{N_\ell} \Delta_t^{i,\ell}\right| \leq N_\ell^{-1}\sum_{i=1}^{N_\ell}\mathbb{E}|\Delta_t^{i,\ell}| \leq c \sum_{\ell'=\ell-1}^{\ell}\left(\mathbb{E}|Y_{\eta_{\ell'}(t)}^{1,n,\ell'}|^p + \mathbb{E}|Y_{\eta_{\ell'}(t)+h_{\ell'}}^{1,n,\ell'}|^p\right).$$

Similarly, we can show that

$$\mathbb{E}\left|\frac{1}{N_0}\sum_{i=1}^{N_0} P_t^{i,0}\right| \leq c\left(\mathbb{E}|Y_{\eta_0(t)}^{1,n,0}|^p + \mathbb{E}|Y_{\eta_0(t)+h_0}^{1,n,0}|^p\right).$$

Note that

$$\begin{aligned}
\mathbb{E}\left|\int_{\mathbb{R}^d}|x|^p\mathcal{M}_t^{(n)}(dx)\right| &\leq \mathbb{E}\left|\frac{1}{N_0}\sum_{i=1}^{N_0}P_t^{i,0} + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\Delta_t^{i,\ell}\right| \\
&\leq c\sum_{\ell=0}^{L}\left(\mathbb{E}|Y_{\eta_\ell(t)}^{1,n,\ell}|^p + \mathbb{E}|Y_{\eta_\ell(t)+h_\ell}^{1,n,\ell}|^p\right).
\end{aligned}$$

We can see from the proof of Lemma 4.1 that the constant c in Lemma 3.2 does not depend on the particular Picard step. Therefore, by Lemma 3.2,

$$\sup_{0\leq t\leq T}\mathbb{E}\left|\int_{\mathbb{R}^d}|x|^p\mathcal{M}_t^{(n)}(dx)\right| \leq c\left(1 + \int_0^T \sup_{0\leq u\leq s}\mathbb{E}\left|\int_{\mathbb{R}^d}|x|^p\mathcal{M}_u^{(n-1)}(dx)\right| ds\right).$$

By iteration, we conclude that

$$\begin{aligned}
\sup_{0\leq t\leq T}\mathbb{E}\left|\int_{\mathbb{R}^d}|x|^p\mathcal{M}_t^{(n)}(dx)\right| &\leq \sum_{r=0}^{n-1}\frac{(cT)^r}{r!} + \sup_{0\leq t\leq T}\mathbb{E}\left|\int_{\mathbb{R}^d}|x|^p\mathcal{M}_t^{(0)}(dx)\right|\frac{(cT)^n}{n!} \\
&\leq e^{cT}\left(1 + \sup_{0\leq t\leq T}\mathbb{E}\left|\int_{\mathbb{R}^d}|x|^p\mathcal{M}_t^{(0)}(dx)\right|\right) < +\infty.
\end{aligned}$$

□

Lemma 4.3 (Verification of **($\mathcal{V}$-Reg)**). *Assume **(Lip)** and **(μ_0-L_p)** . Given any Lipschitz continuous function $C^{0,2}_{b,b} \ni P : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and $n \in \mathbb{N} \cup \{0\}$, there exists a constant c such that*

$$\mathbb{E}\left|\langle \mathcal{M}^{(n)}_t, P(x,\cdot)\rangle - \langle \mathcal{M}^{(n)}_s, P(x,\cdot)\rangle\right|^2 \le c(t-s), \tag{4.1}$$

for any $x \in \mathbb{R}^d$ and $0 \le s \le t \le T$.

Proof. When analysing the regularity of MLMC measure (4.1) one needs to pay attention to the interpolation in time that we used. Pick any $\ell^* \in \{0,1,2,\dots L\}$. For simplicity of notation, we rewrite $\langle \mathcal{M}^{(n)}_t, P(x,\cdot)\rangle$ as

$$\langle \mathcal{M}^{(n)}_t, P(x,\cdot)\rangle := \frac{1}{N_{n,0}} \sum_{i=1}^{N_{n,0}} P^{i,0}_t + \sum_{\ell=1}^{L} \frac{1}{N_{n,\ell}} \sum_{i=1}^{N_{n,\ell}} \left(P^{i,\ell}_t - P^{i,\ell-1}_t\right), \tag{4.2}$$

where

$$P^{i,\ell}_t = \left(\frac{t-\eta_\ell(t)}{h_\ell}\right) P\big(x, Y^{i,n,\ell}_{\eta_\ell(t)+h_\ell}\big) + \left(1 - \frac{t-\eta_\ell(t)}{h_\ell}\right) P\big(x, Y^{i,n,\ell}_{\eta_\ell(t)}\big).$$

Given any $k \in \{0,1,\dots,2^L-1\}$, we compute

$$\begin{aligned}
&\langle \mathcal{M}^{(n)}_{t^{\ell^*}_{k+1}}, P(x,\cdot)\rangle - \langle \mathcal{M}^{(n)}_{t^{\ell^*}_{k}}, P(x,\cdot)\rangle \\
=\ & \frac{1}{N_{n,0}} \sum_{i=1}^{N_{n,0}} \big(P^{i,0}_{t^{\ell^*}_{k+1}} - P^{i,0}_{t^{\ell^*}_{k}}\big) + \sum_{\ell=1}^{L} \frac{1}{N_{n,\ell}} \sum_{i=1}^{N_{n,\ell}} \Big(\big(P^{i,\ell}_{t^{\ell^*}_{k+1}} - P^{i,\ell}_{t^{\ell^*}_{k}}\big) - \big(P^{i,\ell-1}_{t^{\ell^*}_{k+1}} - P^{i,\ell-1}_{t^{\ell^*}_{k}}\big)\Big).
\end{aligned}$$

Thus, we only need to consider $P^{i,\ell}_{t^{\ell^*}_{k+1}} - P^{i,\ell}_{t^{\ell^*}_{k}}$, for each $\ell \in \{0,1,\dots,L\}$. There are two cases depending on the value of ℓ: $\ell < \ell^*$ and $\ell \ge \ell^*$.

For levels $\ell < \ell^*$, at least one of $P^{i,\ell}_{t^{\ell^*}_{k+1}}$ and $P^{i,\ell}_{t^{\ell^*}_{k}}$ is an interpolated value. Then there exist a unique $s \in \{0,1,\dots,2^\ell-1\}$ (chosen such that $\eta_\ell(t^{\ell^*}_k) = t^\ell_s$) and constants $\lambda \in (0, 1-\frac{h_{\ell^*}}{h_\ell}]$ and $\tilde\lambda$, given by

$$\lambda = \frac{t^{\ell^*}_k - t^\ell_s}{h_\ell} \text{ and } \tilde\lambda = \frac{t^{\ell^*}_{k+1} - t^\ell_s}{h_\ell},$$

such that

$$P^{i,\ell}_{t^{\ell^*}_k} = (1-\lambda)P(x, Y^{i,n,\ell}_{t^\ell_s}) + \lambda P(x, Y^{i,n,\ell}_{t^\ell_{s+1}}) \text{ and } P^{i,\ell}_{t^{\ell^*}_{k+1}} = (1-\tilde\lambda)P(x, Y^{i,n,\ell}_{t^\ell_s}) + \tilde\lambda P(x, Y^{i,n,\ell}_{t^\ell_{s+1}}).$$

Note that $\tilde\lambda - \lambda = \dfrac{h_{\ell^*}}{h_\ell}$. By taking the difference between $P^{i,\ell}_{t^{\ell^*}_{k+1}}$ and $P^{i,\ell}_{t^{\ell^*}_{k}}$, we compute that

$$P^{i,\ell}_{t^{\ell^*}_{k+1}} - P^{i,\ell}_{t^{\ell^*}_{k}} = \frac{h_{\ell^*}}{h_\ell}\big(P(x, Y^{i,n,\ell}_{t^\ell_{s+1}}) - P(x, Y^{i,n,\ell}_{t^\ell_s})\big). \tag{4.3}$$

For levels $\ell \geq \ell^*$, both of them are not interpolated. This gives

$$P^{i,\ell}_{t^{\ell^*}_{k+1}} - P^{i,\ell}_{t^{\ell^*}_{k}} = P(x, Y^{i,n,\ell}_{t^{\ell^*}_{k+1}}) - P(x, Y^{i,n,\ell}_{t^{\ell^*}_{k}}). \tag{4.4}$$

By Lemmas 4.2 and 4.1, the hypotheses of Lemma 3.3 are satisfied. By applying Lemma 3.3 to (4.3) and (4.4) along with the global Lipschitz property of P, we have

$$\mathbb{E}|P^{i,\ell}_{t^{\ell^*}_{k+1}} - P^{i,\ell}_{t^{\ell^*}_{k}}|^2 \leq ch_{\ell^*} \quad \forall \ell \in \{0, 1, \ldots, L\}.$$

This shows that

$$\begin{aligned}
& \mathbb{E}\Big|\langle \mathcal{M}^{(n)}_{t^{\ell^*}_{k+1}}, P(x,\cdot)\rangle - \langle \mathcal{M}^{(n)}_{t^{\ell^*}_{k}}, P(x,\cdot)\rangle\Big|^2 \\
\leq \;& \frac{1}{N_{n,0}} \sum_{i=1}^{N_{n,0}} \mathbb{E}|P^{i,0}_{t^{\ell^*}_{k+1}} - P^{i,0}_{t^{\ell^*}_{k}}|^2 + \sum_{\ell=1}^{L} \frac{2}{N_{n,\ell}} \sum_{i=1}^{N_{n,\ell}} \Big(\mathbb{E}|P^{i,\ell}_{t^{\ell^*}_{k+1}} - P^{i,\ell}_{t^{\ell^*}_{k}}|^2 + \mathbb{E}|P^{i,\ell-1}_{t^{\ell^*}_{k+1}} - P^{i,\ell-1}_{t^{\ell^*}_{k}}|^2\Big) \\
\leq \;& ch_{\ell^*}.
\end{aligned}$$

The proof is complete by replacing s and t by $\eta_L(s)$ and $\eta_L(t)$ respectively if any of them (or both) does not belong to Π^L. ☐

Lemma 4.4 below gives a decomposition of MSE (mean-square-error) for MLMC along one iteration of the particle system (1.9).

Lemma 4.4. *Assume **(Ker-Reg)** and **(μ_0-L_p)**. Let $P \in C^{0,2}_{b,b}(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R})$ be a Lipschitz continuous function. Let*

$$MSE^{(m)}_t(P(x,\cdot)) := \mathbb{E}\Big[\Big(\mathbb{E}[P(x, X_t)] - \langle \mathcal{M}^{(m)}_t, P(x,\cdot)\rangle\Big)^2\Big], \quad t \in [0,T].$$

Then, there exists a constant $c > 0$ (independent of the choices of m, L and $(N_{m,\ell})_{0\leq\ell\leq L}$) such that for every $t \in [0,T]$,

$$\begin{aligned}
& \int_{\mathbb{R}^d} MSE^{(m)}_{\eta_L(t)}(P(x,\cdot))\, \mu^{\bar{Z}^L}_{\eta_L(t)}(dx) \\
\leq \;& c\Big(h_L^2 + \int_0^t \Big[\int_{\mathbb{R}^d} \mathbb{E}\Big|\langle \mathcal{M}^{(m-1)}_{\eta_L(s)}, b(x,\cdot)\rangle - \mathbb{E}[b(x, X_{\eta_L(s)})]\Big|^2 \mu^{\bar{Z}^L}_{\eta_L(s)}(dx)\Big] ds \\
& + \int_0^t \Big[\int_{\mathbb{R}^d} \mathbb{E}\Big\|\langle \mathcal{M}^{(m-1)}_{\eta_L(s)}, \sigma(x,\cdot)\rangle - \mathbb{E}[\sigma(x, X_{\eta_L(s)})]\Big\|^2 \mu^{\bar{Z}^L}_{\eta_L(s)}(dx)\Big] ds + \sum_{\ell=0}^{L} \frac{h_\ell}{N_{m,\ell}}\Big).
\end{aligned}$$

Furthermore, if we assume that the functions b and σ are both bounded, then there exists a constant $c > 0$ (independent of the choices of m, L and $(N_{m,\ell})_{0\leq\ell\leq L}$) such that for every

$t \in [0,T]$,

$$\begin{aligned}
&\sup_{x\in\mathbb{R}^d} MSE^{(m)}_{\eta_L(t)}\big(P(x,\cdot)\big)\\
\leq\;& c\bigg(h_L^2 + \int_0^t \bigg[\sup_{x\in\mathbb{R}^d} \mathbb{E}\Big|\langle \mathcal{M}^{(m-1)}_{\eta_L(s)}, b(x,\cdot)\rangle - \mathbb{E}[b(x,X_{\eta_L(s)})]\Big|^2\bigg]\, ds\\
&+\int_0^t \bigg[\sup_{x\in\mathbb{R}^d} \mathbb{E}\Big\|\langle \mathcal{M}^{(m-1)}_{\eta_L(s)}, \sigma(x,\cdot)\rangle - \mathbb{E}[\sigma(x,X_{\eta_L(s)})]\Big\|^2\bigg]\, ds + \sum_{\ell=0}^{L}\frac{h_\ell}{N_{m,\ell}}\bigg).
\end{aligned}$$

Proof. For $x \in \mathbb{R}^d$ and $t \in [0,T]$, we consider

$$\begin{aligned}
&\mathbb{E}\bigg[\Big(\mathbb{E}[P(x,X_{\eta_L(t)})] - \langle \mathcal{M}^{(m)}_{\eta_L(t)}, P(x,\cdot)\rangle\Big)^2\bigg]\\
=\;& \mathbb{E}\bigg[\bigg(\mathbb{E}[P(x,X_{\eta_L(t)})] - \mathbb{E}\Big[\langle \mathcal{M}^{(m)}_{\eta_L(t)}, P(x,\cdot)\rangle\Big|\mathcal{F}^{m-1}\Big]\\
&+\mathbb{E}\Big[\langle \mathcal{M}^{(m)}_{\eta_L(t)}, P(x,\cdot)\rangle\Big|\mathcal{F}^{m-1}\Big] - \langle \mathcal{M}^{(m)}_{\eta_L(t)}, P(x,\cdot)\rangle\bigg)^2\bigg].
\end{aligned}$$

Observe that

$$\begin{aligned}
&MSE^{(m)}_{\eta_L(t)}\big(P(x,\cdot)\big)\\
=\;& \mathbb{E}\bigg[\Big(\mathbb{E}[P(x,X_{\eta_L(t)})] - \mathbb{E}[P(x,Y^{1,m,L}_{\eta_L(t)})|\mathcal{F}^{m-1}]\Big)^2\bigg]\\
&+\mathbb{E}\bigg[\bigg(\mathbb{E}\Big[\langle \mathcal{M}^{(m)}_{\eta_L(t)}, P(x,\cdot)\rangle\Big|\mathcal{F}^{m-1}\Big] - \langle \mathcal{M}^{(m)}_{\eta_L(t)}, P(x,\cdot)\rangle\bigg)^2\bigg],
\end{aligned}\tag{4.5}$$

as $\mathbb{E}\Big[\langle \mathcal{M}^{(m)}_{\eta_L(t)}, P(x,\cdot)\rangle\Big|\mathcal{F}^{m-1}\Big] = \mathbb{E}[P(x,Y^{1,m,L}_{\eta_L(t)})|\mathcal{F}^{m-1}]$ by exchangeability. Next, from Lemma 3.9, there exists a constant c such that

$$\begin{aligned}
&\mathbb{E}\bigg[\Big(\mathbb{E}[P(x,X_{\eta_L(t)})] - \mathbb{E}[P(x,Y^{1,m,L}_{\eta_L(t)})|\mathcal{F}^{m-1}]\Big)^2\bigg]\\
\leq\;& c\,\bigg(h_L^2 + \int_0^t \bigg[\int_{\mathbb{R}^d} \mathbb{E}\Big|\langle \mathcal{M}^{(m-1)}_{\eta_L(s)}, b(x,\cdot)\rangle - \mathbb{E}[b(x,X_{\eta_L(s)})]\Big|^2 \mu^{\bar{Z}^L}_{\eta_L(s)}(dx)\bigg]\, ds\\
&+\int_0^t \bigg[\int_{\mathbb{R}^d} \mathbb{E}\Big\|\langle \mathcal{M}^{(m-1)}_{\eta_L(s)}, \sigma(x,\cdot)\rangle - \mathbb{E}[\sigma(x,X_{\eta_L(s)})]\Big\|^2 \mu^{\bar{Z}^L}_{\eta_L(s)}(dx)\bigg]\, ds\bigg).
\end{aligned}\tag{4.6}$$

By Lemma 3.5, there exists a constant c such that

$$\begin{aligned}&\int_{\mathbb{R}^d}\mathbb{E}\left[\left(\mathbb{E}\left[\langle\mathcal{M}^{(m)}_{\eta_L(t)},P(x,\cdot)\rangle\middle|\mathcal{F}^{m-1}\right]-\langle\mathcal{M}^{(m)}_{\eta_L(t)},P(x,\cdot)\rangle\right)^2\right]\mu^{\bar{Z}^L}_{\eta_L(t)}(dx)\\&=\int_{\mathbb{R}^d}\mathbb{E}\left[\mathrm{Var}\left(\langle\mathcal{M}^{(m)}_{\eta_L(t)},P(x,\cdot)\rangle\middle|\mathcal{F}^{m-1}\right)\right]\mu^{\bar{Z}^L}_{\eta_L(t)}(dx)\le c\sum_{\ell=0}^{L}\frac{h_\ell}{N_{m,\ell}}.\end{aligned}\tag{4.7}$$

Combining (4.5), (4.6) and (4.7) yields the result. □

The complete algorithm consists of a sequence of nested MLMC estimators $\left\{\langle\mathcal{M}^{(m)},P(x,\cdot)\rangle\right\}_{m=1,\ldots,M}$ and its error analysis is presented in Theorem 1.1. Note that we iterate the algorithm by replacing P by the component real-valued functions $\{b_i\}_{1\le i\le d}$ and $\{\sigma_{i,j}\}_{1\le i\le d,1\le j\le r}$.

4.2 Proof of Theorem 1.1

Proof. First, the assumption that $Y^{i,0,\ell}=X_0$ gives

$$\begin{aligned}\sup_{0\le t\le T}\int_{\mathbb{R}^d}\mathbb{E}\Bigg[&\left|\mathbb{E}[b(x,X_{\eta_L(t)})]-\langle\mathcal{M}^{(0)}_{\eta_L(t)},b(x,\cdot)\rangle\right|^2\\&+\left\|\mathbb{E}[\sigma(x,X_{\eta_L(t)})]-\langle\mathcal{M}^{(0)}_{\eta_L(t)},\sigma(x,\cdot)\rangle\right\|^2\Bigg]\mu^{\bar{Z}^L}_{\eta_L(t)}(dx)\le c.\end{aligned}\tag{4.8}$$

Fixing $M>0$ and $P\in C_b^2(\mathbb{R}^d)$, we set

$$a_t^{(m)}:=\begin{cases}\mathbb{E}\left[\left(\langle\mathcal{M}^{(m)}_{\eta_L(t)},P\rangle-\mathbb{E}[P(X_{\eta_L(t)})]\right)^2\right], & m=M,\\ \int_{\mathbb{R}^d}\mathbb{E}\Bigg[\left|\langle\mathcal{M}^{(m-1)}_{\eta_L(t)},b(x,\cdot)\rangle-\mathbb{E}[b(x,X_{\eta_L(t)})]\right|^2 & \\ \qquad+\left\|\langle\mathcal{M}^{(m-1)}_{\eta_L(t)},\sigma(x,\cdot)\rangle-\mathbb{E}[\sigma(x,X_{\eta_L(t)})]\right\|^2\Bigg]\mu^{\bar{Z}^L}_{\eta_L(t)}(dx), & m\le M-1.\end{cases}\tag{4.9}$$

From Lemma 4.4, we observe that

$$a_t^{(m)}\le c\left(b^{(m)}+\int_0^t a_s^{(m-1)}ds\right),\qquad\forall m\in\{1,2,\ldots,M\},\tag{4.10}$$

where $b^{(m)}=h_L^2+\sum_{\ell=0}^{L}\frac{h_\ell}{N_{m,\ell}}$. Then one can easily show that

$$\sup_{0\le t\le T}a_t^M\le\sum_{m=0}^{M-1}b^{(M-m)}\frac{(cT)^m}{m!}+\left(\sup_{0\le s\le T}a_s^{(0)}\right)\cdot\frac{(cT)^M}{M!}.\tag{4.11}$$

Inequalities (4.8) and (4.11) conclude the proof. □

We are now in a position to present the complexity theorem for iterated MLMC estimators of $\{\mathbb{E}[P(X_{\eta_L(t)})]\}_{t\in[0,T]}$.

Theorem 4.5. *Assume* ***(Ker-Reg)*** *and* ***(μ_0-L_p)*** *. Fix $M > 0$ and let $P \in C_b^2(\mathbb{R}^d)$. Then there exists some constant $c > 0$ (independent of the choices of M, L and $\{N_{m,\ell}\}_{m,\ell}$) such that for any $\epsilon < e^{-1}$, there exist M, L and $\{N_{m,\ell}\}_{m,\ell}$ such that for every $t \in [0, T]$,*

$$MSE^{(M)}_{\eta_L(t)}(P) := \mathbb{E}\left[\left(\langle \mathcal{M}^{(M)}_{\eta_L(t)}, P\rangle - \mathbb{E}[P(X_{\eta_L(t)})]\right)^2\right] \le c\,\epsilon^2,$$

and computational complexity is of the order $\epsilon^{-4}|\log\epsilon|^3$.

Proof. The cost of obtaining $\langle \mathcal{M}^{(M)}_{\eta_L(t)}, P\rangle$ involves M iterations. In each iteration, one performs the standard MLMC algorithm, where the cost of approximating the law in the drift and diffusion coefficients is $\sum_{\ell'=0}^{L} N_{m-1,\ell'}$. Hence the overall cost $C := C(M, L, \{N_{m,\ell}\}_{m,\ell})$ of the algorithm is

$$C = \sum_{\ell=0}^{L} h_\ell^{-1} N_{1,\ell} + \sum_{m=2}^{M}\sum_{\ell=0}^{L}\left(h_\ell^{-1} N_{m,\ell}\sum_{\ell'=0}^{L} N_{m-1,\ell'}\right). \tag{4.12}$$

For convenience, we use the notation $x \lesssim y$ to denote that there exists a constant c such that $x \le c\, y$. We shall establish specific values $M^*, L^*, \{N^*_{m,\ell}\}_{m,\ell}$ (depending on ϵ) such that the mean-square error satisfies

$$\sum_{m=1}^{M^*} \frac{c^{M^*-m}}{(M^*-m)!}\left(h_{L^*}^2 + \sum_{\ell=0}^{L^*}\frac{h_\ell}{N^*_{m,\ell}}\right) + \frac{c^{M^*-1}}{M^*!} \lesssim \epsilon^2 \tag{4.13}$$

and show that corresponding computational complexity is of order $\epsilon^{-4}|\log\epsilon|^3$. Firstly, we define

$$M^* := \lfloor \log(\epsilon^{-1})\rfloor \implies c^{M^*-1}(M^*!)^{-1} \lesssim \epsilon^2 \tag{4.14}$$

by Stirling's approximation. For $m \in \{1,\ldots,M^*\}$, we define $\epsilon_m^2 := w_m\epsilon^2$, for some sequence $\{w_m\}_{m=1}^{M^*}$ (depending on M^* and ϵ) which satisfies the following conditions:

(C1) Minimum condition: For each m, $w_m \ge w_{M^*} = 1$;

(C2) Weight condition: $\sum_{m=1}^{M^*}\frac{c^{M^*-m}}{(M^*-m)!} w_m \le K$;

(C3) Cost condition: $\sum_{m=1}^{M^*} w_m^{-1} \le K$,

for some constant $K > 0$. (See Lemma A.2 for a concrete example.) Subsequently, we define

$$L^* := \max_{1\le m\le M^*} L^*_m, \quad L^*_m := \begin{cases} \left|\,\lfloor \log(\epsilon_m^{-1})\rfloor\,\right|, & \epsilon_m \le e, \\ 1, & \epsilon_m > e. \end{cases} \tag{4.15}$$

We also define

$$N_{m,\ell}^* := \left\lceil \epsilon_m^{-2}(L^*+1)h_\ell \right\rceil, \quad \ell \in \{0,\dots,L^*\}, \quad m \in \{1,\dots,M^*\}. \tag{4.16}$$

Note that $h_{L^*} \lesssim \epsilon_m$, for any $m \in \{1,\dots,M^*\}$. To see this, we show that $h_{L_m^*} \lesssim \epsilon_m$ by considering the following three cases.

1. Case I: $\epsilon_m > e$. In this case,
$$h_{L_m^*} = T2^{-L_m^*} = T2^{-1} = \left(\frac{T2^{-1}}{e}\right)e < \left(\frac{T2^{-1}}{e}\right)\epsilon_m.$$

2. Case II: $1 \le \epsilon_m \le e$. In this case,
$$h_{L_m^*} = T2^{-L_m^*} = T2^{\lfloor \log(\epsilon_m^{-1}) \rfloor} = T2^{-\log(\epsilon_m)} \le T \le T\epsilon_m.$$

3. Case III: $0 < \epsilon_m < 1$. Without loss of generality, we assume that $T \le \frac{1}{2}$. (We can scale T by an appropriate factor if it is greater than $\frac{1}{2}$.) In this case,
$$\log(\epsilon_m) \le \left(\frac{1}{\log 2}\right)\log(\epsilon_m) - \frac{\log(2T)}{\log 2} = \frac{\log(\frac{\epsilon_m}{2T})}{\log 2} = \log_2\left(\frac{\epsilon_m}{2T}\right),$$
which implies that
$$h_{L_m^*} = T2^{-L_m^*} = T2^{-\lfloor \log(\epsilon_m^{-1}) \rfloor} \le T2^{-\left(\log(\epsilon_m^{-1})-1\right)} = 2T2^{\log(\epsilon_m)} \le \epsilon_m.$$

We can therefore observe that

$$\begin{aligned}
&\sum_{m=1}^{M^*} \frac{c^{M^*-m}}{(M^*-m)!}\left(h_{L^*}^2 + \sum_{\ell=0}^{L^*} \frac{h_\ell}{N_{m,\ell}^*}\right) \\
\le\ & \sum_{m=1}^{M^*} \frac{c^{M^*-m}}{(M^*-m)!}\left(h_{L^*}^2 + \sum_{\ell=0}^{L^*} \frac{h_\ell}{\epsilon_m^{-2}(L^*+1)h_\ell}\right) \\
\lesssim\ & \sum_{m=1}^{M^*} \frac{c^{M^*-m}}{(M^*-m)!}\epsilon_m^2 \lesssim \epsilon^2,
\end{aligned}$$

by property (C2). Combining this estimate with (4.14), we conclude that the constraint (4.13) is satisfied.

It remains to compute the complexity of the cost under the values $M^*, L^*, \{N^*_{m,\ell}\}_{m,\ell}$.

$$
\begin{aligned}
C &= \sum_{\ell=0}^{L^*} \left(h_\ell^{-1} \left\lceil \epsilon_1^{-2}(L^*+1)h_\ell \right\rceil \right) + \sum_{m=2}^{M^*} \sum_{\ell=0}^{L^*} \Bigg(h_\ell^{-1} \left\lceil \epsilon_m^{-2}(L^*+1)h_\ell \right\rceil \\
&\quad \sum_{\ell'=0}^{L^*} \left\lceil \epsilon_{m-1}^{-2}(L^*+1)h_{\ell'} \right\rceil \Bigg) \\
&\lesssim \sum_{\ell=0}^{L^*} \left(h_\ell^{-1} \left(\epsilon_1^{-2}(L^*+1)h_\ell + 1 \right) \right) + \sum_{m=2}^{M^*} \sum_{\ell=0}^{L^*} \Bigg(h_\ell^{-1} \left(\epsilon_m^{-2}(L^*+1)h_\ell + 1 \right) \\
&\quad \left(\epsilon_{m-1}^{-2}(L^*+1) + (L^*+1) \right) \Bigg) \\
&\lesssim \epsilon^{-2}(L^*+1)^2 + \sum_{m=2}^{M^*} \Bigg(\epsilon_m^{-2}\epsilon_{m-1}^{-2}(L^*+1)^3 + \epsilon_m^{-2}(L^*+1)^3 + \\
&\quad \epsilon^{-1}(L^*+1)^2 \epsilon_{m-1}^{-2} + \epsilon^{-1}(L^*+1)^2 \Bigg) \\
&\lesssim \epsilon^{-2}|\log(\epsilon^{-1})|^2 + |\log(\epsilon^{-1})|^3 \sum_{m=2}^{M^*} \epsilon_m^{-2}\epsilon_{m-1}^{-2} + |\log(\epsilon^{-1})|^3 \sum_{m=2}^{M^*} \epsilon_m^{-2} \\
&\quad + \epsilon^{-1}|\log(\epsilon^{-1})|^2 \sum_{m=2}^{M^*} \epsilon_{m-1}^{-2} + \epsilon^{-1}|\log(\epsilon^{-1})|^2 M^*,
\end{aligned} \tag{4.17}
$$

where, we have used in the last two estimates the bounds $L^* \le \log(\epsilon^{-1})$ (by property (C1)) and $h_\ell^{-1} = T^{-1}2^\ell \le T^{-1}2^{L^*} \lesssim 2^{\log(\epsilon^{-1})} \lesssim \epsilon^{-1}$. Finally, by properties (C1) and (C3) of $\{w_m\}_{m=1}^{M^*}$, together with (4.17) and (4.14), we conclude that $C \lesssim \epsilon^{-4}|\log(\epsilon)|^3$. ☐

4.3 Non-interacting kernels

Here we remark how the theory developed in this work would simplify, if we only treated McKV-SDEs with non-interacting kernels given by

$$
dX_t = b\left(X_t, \int_{\mathbb{R}^d} f(y)\,\mu_t^X(dy) \right) dt + \sigma\left(X_t, \int_{\mathbb{R}^d} g(y)\,\mu_t^X(dy) \right) dW_t, \tag{4.18}
$$

for some continuous functions $b : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^{d \otimes r}$. We assume **(Ker-Reg)** and **(μ_0-L_p)** . We also assume that each component function of f and g belongs to the set $C_b^2(\mathbb{R}^d, \mathbb{R}^q)$. The corresponding MLMC particle system is

$$
dY_t^{i,m,\ell} = b\left(Y_{\eta_\ell(t)}^{i,m,\ell}, \langle \mathcal{M}_{\eta_\ell(t)}^{(m-1)}, f \rangle \right) dt + \sigma\left(Y_{\eta_\ell(t)}^{i,m,\ell}, \langle \mathcal{M}_{\eta_\ell(t)}^{(m-1)}, g \rangle \right) dW_t^{i,m}.
$$

To study this case, we adopt the abstract framework with $\overline{b}(x,\mu) := b(x, \langle \mu, f\rangle)$, $\overline{\sigma}(x,\mu) := \sigma(x, \langle \mu, g\rangle)$ and $\mathcal{V}$ being defined as before. Clearly, this is a special case of the equation studied so far and hence all the results apply. The main difference stems from the complexity analysis as the term $\sum_{\ell=0}^{L} h_\ell^{-1} N_{m,\ell} \sum_{\ell'=0}^{L} N_{m-1,\ell'}$ in (4.12) is replaced by $\sum_{\ell=0}^{L} h_\ell^{-1} N_{m,\ell} +$
$\sum_{\ell'=0}^{L} h_{\ell'}^{-1} N_{m-1,\ell'}$. By performing the same computation as in the proof of Theorem 4.5, we can show that the computational complexity is reduced to the order of $\epsilon^{-2}|\log \epsilon|^2$.

4.4 Plain iterated particle system

The proof of the following theorem constitutes a special case of Lemma 4.4 and Theorem 1.1.

Theorem 4.6. *Assume* ***(Ker-Reg)*** *and* ***(μ_0-L_p)*** *. Fix $M > 0$ and let $P \in C_b^2(\mathbb{R}^d)$. We define the mean-square error as*

$$MSE_t^{(M)}(P) := \mathbb{E}\bigg[\bigg(\frac{1}{N_M}\sum_{i=1}^{N_M} P(\overline{Y}_t^{i,M}) - \mathbb{E}[P(X_t)]\bigg)^2\bigg].$$

Then for every $t \in [0,T]$,

$$MSE_{\eta(t)}^{(M)}(P) \leq c\bigg\{h^2 + \sum_{m=1}^{M} \frac{c^{M-m}}{(M-m)!} \cdot \frac{1}{N_m} + \frac{c^{M-1}}{M!}\bigg\},$$

for some constant $c > 0$ that does not depend on M or $N_1, \dots, N_M$.

The following theorem concerns the computational complexity in the estimation of $\{\mathbb{E}[P(X_{\eta(t)})]\}_{t\in[0,T]}$, whose proof follows similar procedures as the proof of Theorem 4.5 and is omitted.

Theorem 4.7. *Assume* ***(Ker-Reg)*** *and* ***(μ_0-L_p)*** *. Fix $M > 0$ and let $P \in C_b^2(\mathbb{R}^d)$. Then there exists some constant $c > 0$ (independent of the choices of M and $\{N_m\}_{1\leq m\leq M}$) such that for any $\epsilon < e^{-1}$, there exist M and $\{N_m\}_{0\leq m\leq M}$ such that for every $t \in [0,T]$,*

$$MSE_{\eta(t)}^{(M)}(P) := \mathbb{E}\bigg[(\frac{1}{N_M}\sum_{i=1}^{N_M} P(\overline{Y}_{\eta(t)}^{i,M}) - \mathbb{E}[P(X_{\eta(t)})])^2\bigg] \leq c\epsilon^2, \tag{4.19}$$

and computational complexity C is of the order ϵ^{-5}.

5 Numerical results

In this section, we present numerical simulations that confirms that iterative MLMC method achieves one order better computational complexity comparing to classical particle system. Furthermore, numerical experiments indicate that the iterative MLMC method works well even if the coefficients of the McKV-SDEs do not satisfy previously stated regularity and growth assumptions. We compare the following methods

- Classical particle system (1.4),
- MC Picard I - *iterative particle system* (1.8) with fixed number of particles N for all Picard steps,
- MC Picard II - *iterative particle system* (1.8) with an increasing sequence of particles $\{N_m\}_{m=1,\dots,M}$ where $N_m = w_m N_M$ (see the choice of w_m in Lemma A.2),
- Iterated MLMC particle system outlined in Algorithm 1.

5.1 Kuramoto model

First, we provide a numerical example of a one-dimensional stochastic differential equation derived from the Kuramoto model:

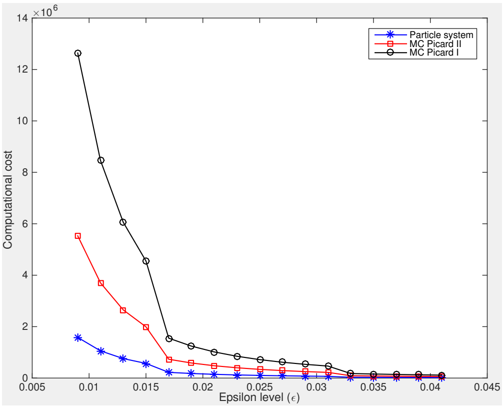
$$\begin{aligned} dX_t &= \int_{\mathbb{R}} \sin(X_t - y)\mu_t^X(dy)\, dt + dW_t, \qquad t \in [0,1], \qquad X_0 = 0, \\ &= \sin(X_t) \int_{\mathbb{R}} \cos(y)\mu_t^X(dy) - \cos(X_t) \int_{\mathbb{R}} \sin(y)\mu_t^X(dy)\, dt + dW_t\,. \end{aligned}$$

For the numerical tests we work with the the bottom representation. We set $P(x) = \sqrt{1+x^2}$. For the initial condition of the iterative algorithm we choose $Y_t^{0,\ell} \sim N(0,t)$.

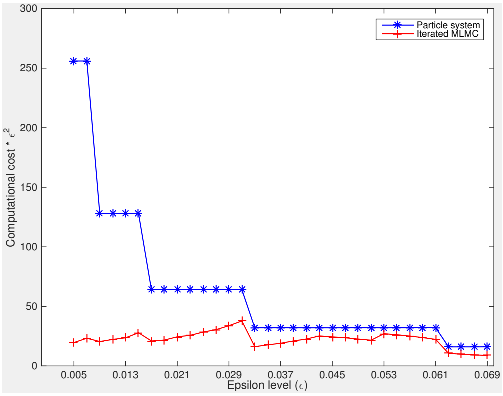
Figure 5.1a shows that both MC Picard I and MC Picard II are less efficient than the classical particle system. In Figure 5.1b, the iterated MLMC particle system achieves computational complexity of order ϵ^{-2} (note that here the cost of simulating particle system is N per Euler step and not N^2 - see Section 4.3).

Figure 5.1c illustrates that the approximation error of iterated methods is within 2ϵ of that of the classical particle system and that it decreases as number of particles increases.

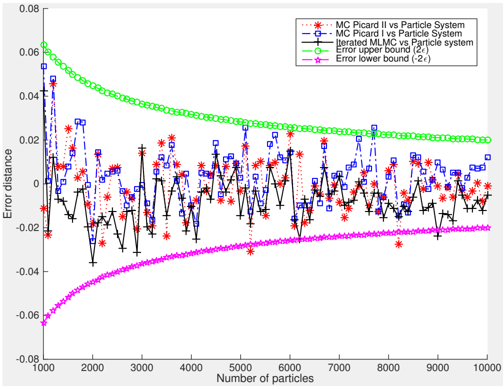
Figure 5.1d depicts $\mathrm{Var}[Y_T^{1,m,\ell} | \mathcal{M}^{(m-1)}]$ and $\mathrm{Var}[Y_T^{1,m,\ell} - Y_T^{1,m,\ell-1} | \mathcal{M}^{(m-1)}]$ (in log scale) for each Picard step across levels ℓ . We see that that the conditional MLMC decays with rate 2. This is higher than the rate given in Lemma 3.4, since this example treats SDE with constant diffusion coefficient for which Euler scheme achieves higher strong convergence rate.



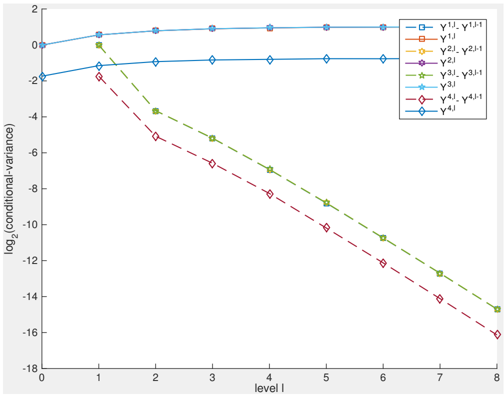
(a) Iterated MC vs Particle method



(b) Iterated MLMC vs Particle method



(c) Approximation error



(d) Variance of $\log_2$ against ℓ for all Picard steps

Figure 5.1: Result of Kuramoto model

5.2 Polynomial drift

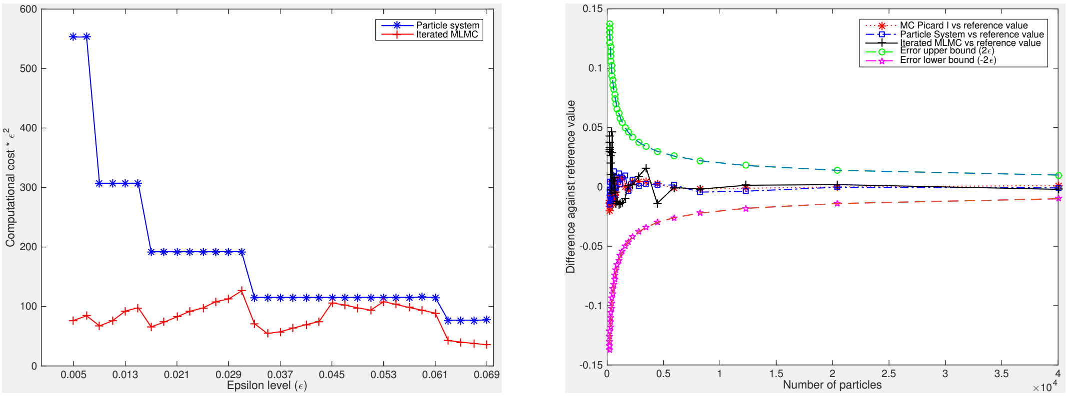
We consider the following McKV-SDE:

$$dX_t = (2X_t + \mathbb{E}[X_t] - X_t\mathbb{E}[X_t^2])dt + X_t dW_t, \quad t \in [0,1], \quad X_0 = 1\,. \tag{5.1}$$

Assumption 2.1 is clearly violated. Note that

$$\begin{aligned} d\mathbb{E}[X_t] &= (3\mathbb{E}[X_t] - \mathbb{E}[X_t]\mathbb{E}[X_t^2])dt \quad \mathbb{E}[X_0] = 1 \\ d\mathbb{E}[X_t^2] &= (5\mathbb{E}[X_t^2] + 2(\mathbb{E}[X_t])^2 - (\mathbb{E}[X_t^2])^2)dt \quad \mathbb{E}[X_0^2] = 1\,. \end{aligned}$$

By solving the above system of ODEs with Euler scheme we obtain particle free approximation to the solution of (5.1) that we use as a reference for iterative MLMC method. Figure 5.2a, shows that the iterated MLMC achieves computational complexity of order ϵ^{-2}. Figure 5.2b indicates that the approximation error of iterated methods is within less than 2ϵ of that of the reference value and that it decreases as number of particles increases.



(a) Iterated MLMC vs Particle sytem (b) Approximation error

Figure 5.2: Result of Polynomial drift

5.3 Viscous Burgers equation

Last, we perform a numerical experiment for the discontinuous case (not Lipschitz) corresponding to the Burgers equation ([4]) given by

$$dX_t = \bar{F}_t(X_t)dt + \frac{1}{4}dW_t, \qquad t \in [0,1], \qquad X_0 = 0, \tag{5.2}$$

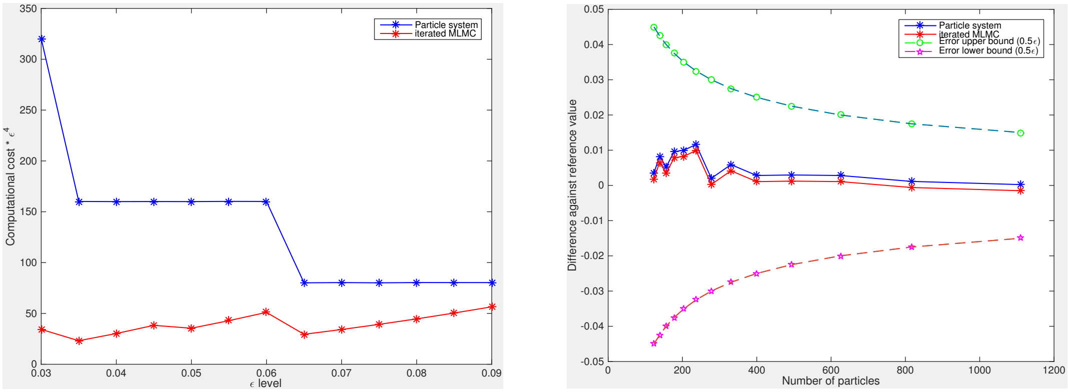
where $\bar{F}_t(x) = \mathbb{P}(X_t \geq x)$. Linking to the Fokker-Planck equation of X_t, it is important to notice that $\bar{F}_t(x)$ is the solution to the viscous Burgers equation:

$$\partial_t v(t,x) = \frac{1}{32}\partial_{xx} v(t,x) - v(t,x)\partial_x v(t,x).$$

where $\bar{F}_0(x) = \mathbf{1}_{\{x \leq 0\}}$ since the initial condition $X_0 = 0$. The Cole-Hopf transformation results in, for any $t \in (0,1]$

$$\bar{F}_t(x) = \frac{\mathcal{N}(\frac{4t-4x}{\sqrt{t}})}{\exp(16x-8t)\mathcal{N}(\frac{4x}{\sqrt{t}}) + \mathcal{N}(\frac{4t-4x}{\sqrt{t}})},$$

where $\mathcal{N}(x) = \int_{-\infty}^{x} \exp(\frac{-y^2}{2})\frac{dy}{\sqrt{2\pi}}$. Then we take $\bar{F}_1(0.5) = 0.5$ as the reference value. In Figure 5.3a, the iterated MLMC achieves computational complexity of order ϵ^{-4}. Figure 5.3b demonstrates the similar desired behaviour of the approximation error as observed in the case of the polynomial drift.



(a) Iterated MLMC vs Particle sytem

(b) Approximation error

Figure 5.3: Result of viscous Burgers equation

A Proofs and useful lemmas

Proof of Lemma 3.2 . Given any ℓ, let us define a sequence of stopping times $\tau_M := \inf\{t \geq 0 : |Z_t^\ell - Z_0^\ell| \geq M\}$. For any $t \in [0,T]$, we consider the stopped process $Z_{t\wedge\tau_M}^\ell$ and compute by the Burkholder-Davis-Gundy and Hölder inequalities and assumptions **($\mathcal{V}$-Lip)** and **(μ_0-L_p)** to obtain that

$$
\begin{aligned}
\mathbb{E}\left[\sup_{0\leq u\leq t} |Z_{u\wedge\tau_M}^\ell|^p\right] &\leq c\left(\mathbb{E}[|Z_0^\ell|^p] + t^{p-1}\mathbb{E}\left[\int_0^t |\overline{b}(Z_{\eta(s)\wedge\tau_M}^\ell, \mathcal{V}_{\eta(s)})|^p ds\right]\right. \\
&\quad \left. + t^{\frac{p}{2}-1}\mathbb{E}\left[\int_0^t \|\overline{\sigma}(Z_{\eta(s)\wedge\tau_M}^\ell, \mathcal{V}_{\eta(s)})\|^p ds\right]\right). \\
&\leq \left(1 + \mathbb{E}\left[\int_0^t \left|\int_{\mathbb{R}^d} |y|^p \mathcal{V}_{\eta(s)}(dy)\right| ds\right]\right. \\
&\quad \left. + \int_0^t \mathbb{E}\left[\sup_{0\leq u\leq s} |Z_{u\wedge\tau_M}^\ell|^p\right] ds\right).
\end{aligned}
$$

Note that, by **(μ_0-L_p)** ,

$$
\mathbb{E}\left[\sup_{0\leq u\leq s} |Z_{u\wedge\tau_M}^\ell|^p\right] \leq c\left(\mathbb{E}\left[\sup_{0\leq u\leq s} |Z_{u\wedge\tau_M}^\ell - Z_0^\ell|^p\right] + \mathbb{E}|Z_0^\ell|^p\right) < +\infty.
$$

By Gronwall's lemma,

$$
\mathbb{E}\left[\sup_{0\leq u\leq t} |Z_{u\wedge\tau_M}^\ell|^p\right] \leq c\left(1 + \mathbb{E}\left[\int_0^T \left|\int_{\mathbb{R}^d} |y|^p \mathcal{V}_{\eta(s)}(dy)\right| ds\right]\right).
$$

Furthermore, since $\sup_{0\leq t\leq T} |Z_{t\wedge\tau_M}^\ell|^p$ is a non-decreasing sequence (in M) converging pointwise to $\sup_{0\leq t\leq T} |Z_t^\ell|^p$, the lemma follows from the monotone convergence theorem.

$\square$

Lemma A.1. *Let $\{Q_t\}_{t\in[0,T]}$ be a cadlag square-integrable process adapted to the filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$. Suppose that $\{W_t\}_{t\in[0,T]}$ is a $\{\mathcal{F}_t\}_{t\in[0,T]}$-Brownian motion. Let $\mathcal{G}$ be a σ-algebra such that $\mathcal{G}\subseteq\mathcal{F}_0$. Then the following equalities hold for any $t\in[0,T]$.*

$$\text{(a)} \qquad \mathbb{E}\Big[\int_0^t Q_s dW_s \Big| \mathcal{G}\Big] = 0,$$
$$\text{(b)} \qquad \mathbb{E}\Big[\Big(\int_0^t Q_s\, dW_s\Big)^2 \Big| \mathcal{G}\Big] = \mathbb{E}\Big[\int_0^t Q_s^2\, ds \Big| \mathcal{G}\Big].$$

The proof follows from standard results of stochastic calculus and is omitted.

Lemma A.2. *The sequence $\{w_m\}_{m=1}^{M^*}$ defined by*

$$w_m := \begin{cases} \max\left\{\frac{(M^*-m-2)!}{c^{M^*-m-2}}, 1\right\}, & 1\le m\le M^*-2, \\ 1, & M^*-1\le m\le M^*, \end{cases}$$

satisfies properties $(C1)$ to $(C3)$ stipulated in the proof of Theorem 4.5.

Proof. First, property (C1) follows easily from the definition of w_m. For property (C2), we verify that

$$\begin{aligned}
&\sum_{m=1}^{M^*} \frac{c^{M^*-m}}{(M^*-m)!} w_m \\
\le\ & \sum_{m=1}^{M^*-2} \frac{c^{M^*-m}}{(M^*-m)!}\Big(\frac{(M^*-m-2)!}{c^{M^*-m-2}} + 1\Big) + \sum_{m=M^*-1}^{M^*} \frac{c^{M^*-m}}{(M^*-m)!} \\
=\ & \sum_{m=1}^{M^*} \frac{c^{M^*-m}}{(M^*-m)!} + c^2 \sum_{m=1}^{M^*-2} \frac{1}{(M^*-m)(M^*-m-1)} \\
=\ & \sum_{m=1}^{M^*} \frac{c^{M^*-m}}{(M^*-m)!} + c^2\Big(1 - \frac{1}{M^*-1}\Big) \le e^c + c^2.
\end{aligned}$$

Lastly, we show this sequence satisfies property (C3). Indeed,

$$\sum_{m=1}^{M^*} w_m^{-1} = \sum_{m=1}^{M^*-2} \frac{c^{M^*-m-2}}{(M^*-m-2)!} + 2 \le e^c + 2.$$

$\square$

Algorithm 1: Nested MLMC with Picard scheme

Input: Initial measure μ^0 for $Y^{i,0,\ell}$, global Lipschitz payoff function $C_p^2 \ni P : \mathbb{R}^d \to \mathbb{R}$ and accuracy level ϵ

Output: $\langle \mathcal{M}_T^{(M)}, P \rangle$, the approximation for our goal $\mathbb{E}[P(X_T)]$.

Fix parameters M (see (4.14)) and L (see (4.15)) that correspond to ϵ;

Given $\mu^0 = \text{Law}(Y^{i,0,0})$, sample $\{Y^{i,0,0}_{t_k^L}\}_{k=0,\dots,2^L}$;

for $m = 1$ *to* $M-1$ **do**

During mth Picard step, given samples $\{Y^{i,m-1,\ell}_{t_k^\ell}\}^{\ell=0,\dots,L}_{k=0,\dots,2^\ell}$, take (1.9) and run MLMC to obtain $\{Y^{i,m,\ell}_{t_k^\ell}\}^{\ell=0,\dots,L}_{k=0,\dots,2^\ell}$. This requires calculating

$$\left(\langle \mathcal{M}^{(m-1)}_{t_0^L}, b(x,\cdot) \rangle, \dots, \langle \mathcal{M}^{(m-1)}_{t_{2^L}^L}, b(x,\cdot) \rangle \right),$$

$$\left(\langle \mathcal{M}^{(m-1)}_{t_0^L}, \sigma(x,\cdot) \rangle, \dots, \langle \mathcal{M}^{(m-1)}_{t_{2^L}^L}, \sigma(x,\cdot) \rangle \right),$$

where in place of x, we put particles $\{Y^{i,m,\ell}_{t_k^\ell}\}^{\ell=0,\dots,L}_{k=0,\dots,2^\ell-1}$;

Given samples $\{Y^{i,M-1,\ell}_{t_k^\ell}\}^{\ell=0,\dots,L}_{k=0,\dots,2^\ell}$, run standard MLMC (with interpolation) to obtain the final vector of approximations $\left(\langle \mathcal{M}^{(M)}_{t_0^L}, P \rangle, \dots, \langle \mathcal{M}^{(M)}_{t_{2^L}^L}, P \rangle \right)$;

Return $\langle \mathcal{M}_T^{(M)}, P \rangle$.

B Algorithm for the MLMC particle system

Acknowledgements

We are grateful to Mireille Bossy, Mike Giles and David Šiška for helpful comments.

References

[1] A. L. H. Ali. *Pedestrian Flow in the Mean Field Limit*. PhD thesis, King Abdullah University of Science and Technology (KAUST), 2012.

[2] F. Antonelli and A. Kohatsu-Higa. Rate of convergence of a particle method to the solution of the McKean–Vlasov equation. *The Annals of Applied Probability*, 12(2):423–476, 2002.

[3] M. Bossy. Optimal rate of convergence of a stochastic particle method to solutions of 1D viscous scalar conservation laws. *Mathematics of computation*, 73(246):777–812, 2004.

[4] M. Bossy, L. Fezoui, and S. Piperno. Comparison of a stochastic particle method and a finite volume deterministic method applied to Burgers equation. *Monte Carlo Methods and Applications*, 3:113–140, 1997.

[5] M. Bossy and B. Jourdain. Rate of convergeance of a particle method for the solution of a 1D viscous scalar conservation law in a bounded interval. *The Annals of Probability*, 30(4):1797–1832, 2002.

[6] M. Bossy and D. Talay. Convergence rate for the approximation of the limit law of weakly interacting particles: application to the Burgers equation. *The Annals of Applied Probability*, 6(3):818–861, 1996.

[7] M. Bossy and D. Talay. A stochastic particle method for the McKean-Vlasov and the Burgers equation. *Mathematics of Computation of the American Mathematical Society*, 66(217):157–192, 1997.

[8] R. Buckdahn, J. Li, S. Peng, and C. Rainer. Mean-field stochastic differential equations and associated pdes. *The Annals of Probability*, 45(2):824–878, 2017.

[9] K. Bujok, B. Hambly, and C. Reisinger. Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives. *Methodology and Computing in Applied Probability*, pages 1–26, 2013.

[10] R. Carmona, F. Delarue, and A. Lachapelle. Control of McKean–Vlasov dynamics versus mean field games. *Mathematics and Financial Economics*, 7(2):131–166, 2013.

[11] J.-F. Chassagneux, D. Crisan, and F. Delarue. A probabilistic approach to classical solutions of the master equation for large population equilibria.

[12] F. Delarue, J. Inglis, S. Rubenthaler, and E. Tanré. Global solvability of a networked integrate-and-fire model of Mckean–Vlasov type. *The Annals of Applied Probability*, 25(4):2096–2133, 2015.

[13] F. Delarue, J. Inglis, S. Rubenthaler, and E. Tanré. Particle systems with a singular mean-field self-excitation. Application to neuronal networks. *Stochastic Processes and their Applications*, 125(6):2451–2492, 2015.

[14] W. E, M. Hutzenthaler, A. Jentzen, and T. Kruse. On full history recursive multilevel picard approximations and numerical approximations of high-dimensional nonlinear parabolic partial differential equations. *arXiv preprint arXiv:1607.03295*, 2016.

[15] A. Friedman. *Stochastic differential equations and applications*. Courier Corporation, 2006.

[16] M. B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[17] M. B. Giles, T. Nagapetyan, and K. Ritter. Multilevel Monte Carlo approximation of distribution functions and densities. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):267–295, 2015.

[18] A.-L. Haji-Ali and R. Tempone. Multilevel and Multi-index Monte Carlo methods for mckean-vlasov equations. *arXiv preprint arXiv:1610.09934*, 2016.

[19] S. Heinrich. Multilevel Monte Carlo methods. In *Large-scale scientific computing*, pages 58–67. Springer, 2001.

[20] A. Kebaier. Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing. *The Annals of Applied Probability*, 15(4):2681–2705, 2005.

[21] N. V. Krylov. *Controlled diffusion processes*, volume 14 of *Applications of Mathematics*. Springer-Verlag, New York-Berlin, 1980. Translated from the Russian by A. B. Aries.

[22] N. V. Krylov. *Introduction to the theory of random processes*, volume 43. American Mathematical Soc., 2002.

[23] V. Lemaire and G. Pagès. Multilevel Richardson–Romberg extrapolation. *Bernoulli*, 23(4A):2643–2692, 2017.

[24] P. Lions. Cours au collège de france: Théorie des jeux à champs moyens, 2014.

[25] H. McKean Jr. A class of Markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences of the United States of America*, 56(6):1907, 1966.

[26] S. Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations*, pages 42–95. Springer, 1996.

[27] K. R. Parthasarathy. *Probability measures on metric spaces*, volume 352. American Mathematical Soc., 1967.

[28] S. B. Pope. *Turbulent flows*. Cambridge University Press, 2000.

[29] L. Ricketson. A multilevel Monte Carlo method for a class of Mckean-Vlasov processes. *arXiv preprint arXiv:1508.02299*, 2015.

[30] A.-S. Sznitman. *Topics in propagation of chaos*. Springer, 1991.

[31] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic analysis and applications*, 8(4):483–509, 1990.